ICTIMES 2019

5th INTERNATIONAL CONFERENCE ON

Trends in Information, Management, Engineering and Sciences

Proceedings of International Conference on Latest Trends in Electronics and Communication (ICLTEC)

ISBN No: 978-93-88808-62-0

Convenor
Dr. R. P. Naik
HOD-Dept. of ECE

Editor
Dr. B BALAJI
Professor - ECE

Editor in Chief
Dr. P. John Paul
Principal

Organized by

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTUH, Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.

website: www.mrce.in

Proceedings on International Conference on Electronics and Communication (ICLTEC-2019)
ISBN 978-93-88808-62-0

ICTIMES - INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION (ICLTEC-2019)

S. No	Title of the Paper	Page No.
IC19EC1	Inventions , Innovations And Startups In Higher Educational Institutions *--Dr.John Paul Pulipati, Dr.Kezia Rani Burgula*	1
IC19EC2	Low Overhead Warning Flip-Flop Based On Charge Sharing For Timing Slack Monitoring *--Pochana Ashok Kumar*	17
IC19EC3	Clock-Gating Of Streaming Applications For Energy Efficient Implementations On FPGA's *--K. Mounika*	27
IC19EC4	Transition Inversion Based Low Power Data Coding Scheme For Buffered Data Transfer *--Suhasini Parunandi*	32
IC19EC5	An Iot Based Solar Vehicle With Obstacle Avoidance Night Speed Limitter And Accident Detection Using GSM And GPS *--Mamidi Prudhvi Raj*	38
IC19EC6	DLAU: A Scalable Deep Learning Accelerator Unit On FPGA *--Vajja Parmesh*	48
IC19EC7	Division Circuit Using Reversible Logic Gates *--Rathnavath Priyadarshini*	53
IC19EC8	Design Of An Energy Efficient Iot Enabled Smart System Based On DALI Network Over MQTT Protocol *--K. Premalatha*	59
IC19EC9	Verification And Simulation Of New Designed NAND Flash Memory Controller *--Akarapu Natraj Renu*	64
IC19EC10	Advanced Video Coding/H.264 Encoding And Decoding Process With Ipppp Sequences *--Anitha Perla, Chella Moses*	69
IC19EC11	Low Power Viterbi Decoder Design Based On Reversible Logic Gates *--Nenavath Naresh*	76
IC19EC12	Design Of High Speed Sequence Detector Using Verilog *--Bejagam Divya*	81
IC19EC13	A Low Power Large Bandwidth Flipped Voltage Follower Based Current Mirror *--P.Anil Kumar, D.Rajesh*	85
IC19EC14	The Gas Leak Detection Based on a Wireless Monitoring System *--K.Suhasini*	90
IC19EC15	Influence Of Wind Power Plants On Power Systems Operation *--Poushya Kandukuri, DRK Mahesh*	102

ISBN 978-93-88808-62-0

IC19EC16	Design & Implementation of Efficient Multiplier Using Fixed Width RPR --*M. Shiva Kumar, A. Suresh Reddy*	107
IC19EC17	Modern Technology in Solar Energy Generation --Nimmala Mahesh,P. Manikya rao	111
IC19EC18	Talking without Talking: A Solution to Noisy Communication K MANIKANTA,M.RANI	117
IC19EC19	Delay Estimation Model for High speed Interconnects in VLSI B.Balaji, R. Purushottam Naik	121
IC19EC20	IOT Based Light Intensity Monitoring System using Embedded Linux & Raspberry Pi --KESANAPALLI ASKA RANI1 ,BODDU MANJULA	127
IC19EC21	Smart Waste Collection Monitoring and Alert System via IoT Vaddepally Anusha	133
IC19EC22	Chip Design for Turbo Encoder Module for In-Vehicle System PULUKURI NEIL SAMUEL SADANAND	139
IC19EC23	Review on Energy Harvesting For Wireless Sensor Networks P.Venkatapathi, Bala tejavath, Godugu Kiran Kumar	144
IC19EC24	A Simple Yet Efficient Accuracy Configurable Adder Design --*Vavaldas Harshini*	148
IC19EC25	Carbon Nanotube Using Electronics --*Kukatla Rajeshwar,Singamaneni Nagajyothi*	162

INVENTIONS , INNOVATIONS AND STARTUPS IN HIGHER EDUCATIONAL INSTITUTIONS

Dr.John Paul Pulipati,Principal ,principal@mrce.in,Malla Reddy College of Engineering,Hyderabad,India

Dr.Kezia Rani Burgula,Associate Professor,CSE Department,keziarani@stanley.edu.in,Stanley College of Engineering for Women,Hyderabad,India

ABSTRACT

The paper advances current knowledge on factors affecting higher education institutions in their quest for innovation in education. This paper takes an innovative approach by adopting the concept of 'innovation systems' and adapting it for higher education. The analytical construct of 'higher education innovation system' has thus been developed as a sub-set of an innovation system, concentrated particularly in higher education institutions which are close connection with other institutional spheres, such as industry, government and non-government agencies, and the society at large in terms of Internship and innovation process. As the part of the interns discussed with classification internship and also how it has been played major role in the institution towards to the Higher Education process. In higher education innovation system has to be following with set of functions, components and relationships, which allow us to disaggregate the various levels of interactions among the elements of the system and analyse the unfolding of innovation in higher education. Results indicate certain "disengagement" in relation of higher education institutions and education policy makers, business, and students as well as between higher education institutions' managers and their subordinates. Based on the findings, major innovation, incubation and startup related challenges in the higher education are discussed and related practical recommendations are presented.

Key words: Innovation, start up, intellectual property rights, entrepreneurship, technology transfer, technology commercialization, technology policy, university-industry links, economic development, R&D, knowledge management, incubation

1. INTRODUCTION

Higher education is changing across Indian and other global countries and there is a growing expectation from policy makers and society that higher education institutions (HEIs) should

evolve into a new type of economic actor. Entrepreneurship and innovation in higher education are no longer only associated with business start-ups and technology transfer but are increasingly understood as core elements of a procedural framework for how organisations and individuals behave. For example, in how links between teaching and research are created and nurtured, how societal engagement and knowledge exchange are organized, how resources are built and managed for effective partnerships, and how new entrepreneurs are supported [1].

One of the major contributions in fields of application and impact of innovation, incubation and internship in HEIs has been laid forward by William Spady, a sociologist and the father of Outcome-Based Education (OBE). OBE is referred to by over 20 different names including Systemic education restructuring, Performance Based Education, Standards based education reform, High Performance Learning, Total Quality Management, Transformational Education, and Competency-Based Education [8-14].Outcome-Based Education[2-6] means clearly focusing and organizing everything in an educational system around what is essential for all students to be able to do successfully at the end of their learning experiences. This means starting with a clear picture of what is important for students to be able to do, then organizing curriculum, instruction, and assessment to make sure this learning ultimately happens. The keys to having an outcome-based system are [9-10]:

1. Developing a cleat set of learning outcomes around which all of the system's components can be focused.
2. Establishing the conditions and opportunities within the system that enable and encourage all students to achieve those essential outcomes.

For example, the possible outcome "explain the major causes of inflation in capitalist economies" implies that to be successful, the learner will be expected to develop both the competence of explaining and knowledge of major causes of inflation in capitalist economies. Since outcome-based systems expect learners to earn out the processes defined within an outcome statement, they are careful to build those processes directly into the outcome through demonstration verbs. Therefore, one key to recognizing a well-defined outcome is to look for the demonstration verb or verbs that define which processes the learner is expected to carry out at the end. Without those verbs, what are called outcome statements lack a clearly defined demonstration process, and without that defined process the outcome statement takes on the character of a goal rather than a true outcome demonstration [11-12].

1.2 Program Educational Objectives (PEO)

Graduates are able to:

1.2.1 PEO-1: Work in automotive and related industries by applying the knowledge of science, mathematics and engineering.

1.2.2 PEO-2 Demonstrate team work and project management skills in a multi-disciplinary environment through effective communication, modern tool usage and professional ethics.

1.2.3 PEO-3 Investigate, analyse, research and solve problems in the field of automotive planning, designing, manufacturing, testing and servicing.
1.2.4 PEO-4 Pursue lifelong learning through professional trainings and practices with consideration of safety, environment and sustainability

Following are the Program Learning Outcomes (PLOs) adopted in Automotive Engineering Program as recommended by PEC in their Manual of Accreditation 2014.

1.2.5 PLO-1 Engineering Knowledge: An ability to apply knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems.
1.2.6 PLO-2 Problem Analysis: An ability to identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences and engineering sciences.
1.2.7 PLO-3 Design/Development of Solutions: An ability to design solutions for complex engineering problems and design systems, components or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.[3,4]
1.2.8 PLO-4 Investigation: An ability to investigate complex engineering problems in a methodical way including literature survey, design and conduct of experiments, analysis and interpretation of experimental data, and synthesis of information to derive valid conclusions.
1.2.9 PLO-5 Modern Tool Usage: An ability to create, select and apply appropriate techniques, resources, and modern engineering and IT tools, including prediction and modeling, to complex engineering activities, with an understanding of the limitations.

1.32.10 PLO-6 The Engineer and Society: An ability to apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the responsibilities relevant to professional engineering practice and solution to complex engineering problems.
1.2.11 PLO-7 Environment and Sustainability: An ability to understand the impact of professional engineering solutions in societal and environmental contexts and demonstrate knowledge of and need for sustainable development.
1.2.12 PLO-8 Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice.
1.2.13 PLO-9 Individual and Team Work: An ability to work effectively, as an individual or in a team, on multifaceted and /or multidisciplinary settings.
1.2.14 PLO-10 Communication: An ability to communicate effectively, orally as well as in writing, on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
1.2.15 PLO-11 Project Management: An ability to demonstrate management skills and apply engineering principles to one's own work, as a member and/or leader in a team, to manage projects in a multidisciplinary environment.

1.2.16 PLO-12 Lifelong Learning: An ability to recognize importance of, and pursue lifelong learning in the broader context of innovation and technological developments.

1.3 A S K (Attitude Skill Knowledge)

Attitude, skill and knowledge[1,2], all the three elements put together form a success formula. It is a well-known fact that any person with a remarkable blend of all the above mentioned 3 characteristics will come out as a human being with lots of intelligence, leadership qualities and mentoring abilities. But, the question is, Can any one of these attributes stand alone and guide a person to the zenith? The answer is probably NO. As, we all know and understand that in the present era of competitiveness and naivety every individual is expected to be skillful, knowledgeable and also must possess a correct set of attitude which is universally acceptable. A person with only a bucket full of skills required to complete a task with a very less amount of knowledge on the subject and probably with a poor attitude cannot be considered useful for an organization for the long run. If you don't have skills, it can be acquired. If you don't have knowledge, it can be gained. But, if you don't have attitude, you are in trouble. Attitude, in a general sense is taken negatively. The triangle of success represented by ASK is shown in Fig. 1 [21-23].

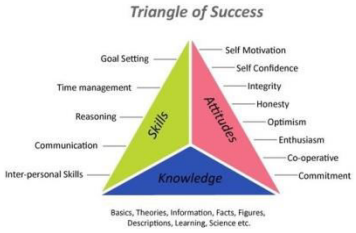


Fig. 1 Attitude, skill and knowledge- The triangle of success

1.4 Blooms Taxonomy

Bloom's taxonomy is a set of three hierarchical models used to classify educational learning objectives into levels of complexity and specificity. The three lists cover the learning objectives in cognitive, affective and sensory domains. Various stages of Blooms taxonomy is shown in Fig. 3 [13-18].

1.3.2 The Revised Taxonomy (2001)

A group of cognitive psychologists, curriculum theorists and instructional researchers, and testing and assessment specialists published in 2001 a revision of Bloom's Taxonomy with the title *A Taxonomy for Teaching, Learning, and Assessment*. This title draws attention away from the somewhat static notion of "educational objectives" (in Bloom's original title) and points to a more dynamic conception of classification.

- Remember
  - Recognizing
  - Recalling
- Understand
  - Interpreting
  - Exemplifying
  - Classifying
  - Summarizing

- Inferring
- Comparing
- Explaining

- Apply
  - Executing
  - Implementing
- Analyze
  - Differentiating
  - Organizing
  - Attributing
- Evaluate
  - Checking
  - Critiquing
- Create
  - Generating
  - Planning
  - Producing

Bloom's revised taxonomy is shown in Fig. 2.

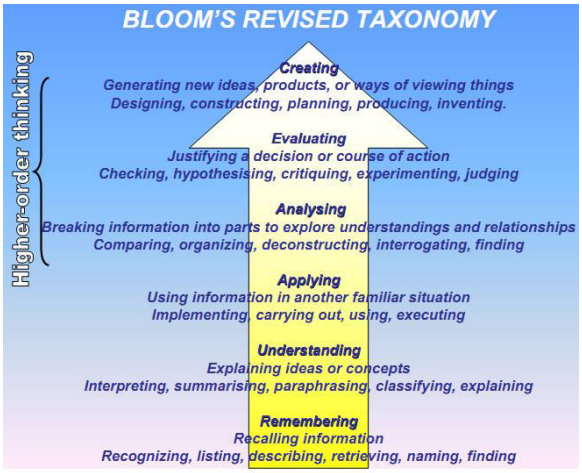


Fig. 2 Bloom's revised taxonomy

II Internship:

An internship is an official program offered by an employer to potential employees. Interns work either part time or full time at a company for a certain period of time. Internships are most popular with undergraduates or graduate students who work between one to four months and have a goal to gain practical work or research related experience. The main difference between an apprenticeship and an internship is that internships are more exploratory. You're not bound to work for your employer after the internship is over (although many interns do receive job offers). If you start early enough to do a few internships throughout college, you can use the first ones to get a feel for what career you'd like to pursue and the later ones to build your experience. Internships can be paid or unpaid. For example, in California, unpaid interns must receive college credit for their work.

2.1 Benefits for Internship Program

1. Tech-savvy. The first generation to be brought up with computers, members of Gen Y are "digital natives." In fact, not only can they uncover, operate, and recommend the most advanced

technologies; they can teach you how to use tools like content management systems and social media.

2. Cost-effective. Compared to other populations, Generation Y appears less motivated by money. A study by UNC's Kenan-Flagler Business School[4] found that millennial prioritize meaningful work over higher pay.

3. Intrinsically motivated. As the results of the Kenan-Flagler Business School study demonstrate, millennial want to grow, develop, and advance within the workplace. They are intrinsically motivated to succeed, a quality all employers seek in their employees.

4. Team players. If some called Generation X "The Me Generation," we might term Generation Y "The We Generation" for their heightened sense of community and peer-to-peer relationships.

5. Highly educated. Millennials are on track to be the most educated generation yet. According to a 2016 study by the Pew Research Center, 20 percent of millennials were college graduates, with an additional 40 percent still in school. About half of millennials still in school planned to earn a graduate or professional school degree, and many millennials who had graduated planned to return.

6. Optimistic. Positivity in the workplace is invaluable to company culture and team morale, just as negativity can be extremely detrimental.

7. Current. Like employing any youth generation, hiring millennials helps keep your company up to date with social, entertainment, and other market trends. Millennials offer a fresh perspective and can generate marketing strategies that appeal to younger generations of consumers.

2.3 Role of Internship coordinator—this person generally works in the host organization's HR department and can be seen as the overall director of the internship program. The internship coordinator is usually responsible for overseeing the development and implementation of an internship program, overseeing and coordinating the program once it's up and running, assigning interns to various departments, and managing site supervisors. This individual also acts as the liaison between the company and the educational institution (the student's faculty sponsor, the school's career centre director, and/or career counselors).

In general, there are three aspects necessary for an internship to constitute a learning experience:

1. The internship is within the student's area of study; and they bring to the internship the knowledge they've acquired through their academic education.
2. The internship supervisor provides guidance, evaluation, and feedback to facilitate the learning process. (Evaluations and feedback, however, can be written or verbal.)
3. The student engages in ongoing contemplation of learning objectives throughout the course of the internship.

2.4 INTERNATIONAL INTERNSHIP: An International Development internship with Projects Abroad and gain skills that you will use for the rest of your career, regardless of what field you move into. As you get hands-on experience in the development field you'll learn about infrastructure and disaster management as well as aid and development in disadvantaged communities. You'll also gain a deeper understanding of the world and the complex processes that are involved in supporting people and places to grow and prosper. We currently have a number of International Development Projects, working to transform communities that need help. Your work could include creating campaigns, educating local communities, and facilitating existing projects. You'll work with local experts and Projects Abroad staff to better understand global development challenges and the work needed to overcome them. There are two main aims to this project: providing you with a practical, insightful work experience, while helping to address issues affecting development in Abroad (Developed Country).Many of the interns who join us are looking for hands-on experience in their field of study. We provide you with the opportunity to learn from skilled local professionals and gain first-hand insight into the issues facing. By sharing knowledge and experiences, you'll participate in valuable cross-cultural exchange. This will help prepare you for your future career. ***The goal of international development is to improve the wellbeing of local communities by working on sustainable solutions to their problems***. On this internship, you can choose to focus on a range of different issues, from immigration to human rights. Working with local NGOs who have expert knowledge of the situation, you'll contribute towards their efforts in protecting and empowering vulnerable groups of people. Help to create long-term positive change in local communities on this internship, while gaining invaluable international development work experience in Mexico.

3. Business incubator: is a company that helps new and startup companies to develop by providing services such as management training or office space. The National Business Incubation Association (NBIA) defines business incubators as a catalyst tool for either regional or national economic development. NBIA categorizes their members' incubators by the following five incubator types: academic institutions; non-profit development corporations; for-profit property development ventures; venture capital firms, and combination of the above [1-2].Most research and technology parks do not offer business assistance services, which are the hallmark of a business incubation program. However, many research and technology parks house incubation programs [3].Incubators also differ from the U.S. Small Business Administration's Small Business Development Centers (and similar business support programs) in that they serve only selected clients. SBDCs are required by law to offer general business assistance to any company that contacts them for help [4-5]. Within European Union countries

there are different EU and state funded programs that offer support in form of consulting, mentoring, prototype creation and other services and co-funding for them. TecHub is one of examples for IT companies and ideas [6].In India, the business incubators are promoted in a varied fashion: as Technology Business Incubators (TBI) and as Startup Incubators -- the first deals with technology business (mostly, consultancy and promoting technology related businesses) and the later deals with promoting startups (with more emphasis on establishing new companies, scaling the businesses, prototyping, patenting, and so forth).

Types of services

Since startup companies lack many resources, experience and networks, incubators provide services which helps them get through initial hurdles in starting up a business. These hurdles include space, funding, legal, accounting, computer services and other prerequisites to running the business [17-20].

Among the most common incubator services are:

Help with business basics

- Networking activities
- Marketing assistance
- Market Research
- Access to angel investors or venture capital
- Comprehensive business training programs
- Advisory boards and mentors
- Management team identification
- Help with business etiquette
- Technology commercialization assistance
- Intellectual property management

Types of Business Incubation :
There are a number of business incubators that have focused on particular industries or on a particular business model, earning them their own name.
This list is incomplete; you can help by expanding it.

- Virtual business incubator - online business incubator
- Kitchen incubator - a business incubator focused on the food industry
- Public incubator - a business incubator focused on the public good
- Seed accelerator - a business incubator focused on early startups

Corporate accelerator - a program of a larger company that acts akin to a seed accelerator

- Startup studio - a business incubator with interacting portfolio companies
- Technology Business Incubator - a business incubator based out academic institutions

- Hybrid Incubator - A business incubator that combines virtual incubator with on-premise activities

A *startup* is a young company founded by one or more entrepreneurs in order to develop a unique product or service and bring it to market. By its nature, the typical *startup* ends to be a shoestring operation, with initial funding from the founders or their families.

A startup venture could be defined as a new business that is in the initial stages of operation, beginning to grow and is typically financed by an individual or small group of individuals. It is a young entrepreneurial, scalable business model built on technology and innovation wherein the founders develop a product or service for which they foresee demand through disruption of existing or by creating entirely new markets. Startups[25-28] are nothing but an idea that manifests into a commercial undertaking.

Grant Thornton (2016) define startup business[26,28] as an organization which is

a) Incorporated for three years or less
a) At a funding stage of Series B or less(B Series means second round of funding)
b) An entrepreneurial venture/a partnership or a temporary business organisation
c) Engages in development, production or distribution of new products/services or processes
d) Revenue of up to INR 25 cr.
e) Not formed through splitting or restructuring
f) Employing 50 people or less

Department of Industrial Policy and Promotion (DIPP) define a startup as an entity incorporated or registered in India with following parameters:

- Established not prior to seven years, (for Biotechnology Startups not prior to ten years)
- With annual turnover not exceeding INR 25Cr in any preceding financial year, and
- Working towards innovation, development or improvement of products or processes or services,
- It is a scalable business model with a high potential of employment generation or wealth creation

It is to be noted that such entity is not formed by splitting up, or reconstruction, of a business already in existence. Also, an entity shall cease to be a startup if its turnover for the previous financial years has exceeded INR 25cr or it has completed 7 years (biotechnology startups 10 years) from the date of incorporation/ registration('Startup India', 2017)

GOVERNMENT INITIATIVES

Indian government is serious in promoting entrepreneurship at the startup level and has taken a number of initiatives to ensure appropriate support. In this aspect it is relevant to mention **'Make in India'** campaign introduced in September'14 to attract foreign investments and encourage domestic companies to participate in the manufacturing sector. The government increased the foreign direct investment (FDI) limits for most of the sectors and strengthened intellectual property rights (IPRs) protection to instill confidence in the startups.In order to make the country as number one destination for startups, Government of India (GoI) has introduced a new campaign called '**Standup India'** in 2015 aimed at promoting entrepreneurship among women and to help startups with bank funding. Another commendable and far reaching initiative is '**Digital India'** introduced in 2015 to ensure government services are made available to every citizen through online platform that aims to connect rural areas by developing their digital infrastructure which translates into a huge business opportunity for startups.

THE STARTUP SCENARIO IN INDIA

It is to be noted that every year more than 800 technology startups are being set up in India. By 2020, it is estimated that around 11,500 tech-startups are going to be established with employment potential of around 250,000 technical people (NASSCOM, 2015). It is admirable to note that India is amongst the top five countries in the world in terms of startups with 10,000+ led by US with 83,000+ comprising 43% tech-based firms with 9% managed by women entrepreneurs. The number of incubators also has crossed 100 in 2014-15 to give boost to the startup saga (Grant Thornton, 2015).Sector wise, the distribution of Indian businesses is:

Table:1 Break-up of Indian Startup Businesses

Technology Based	Non-Technology Based
E-Commerce - 33%	Engineering- 17%
B2B - 24%	Construction-13%
Internet - 12%	Agri- products- 11%
Mobile apps - 10%	Textile - 8%
SaaS - 8%	Printing & packaging – 8%
Other – 13%	Transport & logistics- 6%
	Outsourcing & support -5%
	Others-32%

Source: Startups India- An Overview, Grant Thornton, 2015

THE STARTUP ECOSYSTEM

Along with government initiatives, there is a definite movement in startup arena in India due to penetration of IT and internet. Many startups[30-34] are coming up in service sector including education, legal, retail, insurance and health. With customers becoming aware of the benefits and convenience, the popularity and viability of startups is no more a difficult proposition for an entrepreneur.

A number of venture capitalists and angel investors are aggressive and gung-ho on Indian startups as they see lot of potential with few expected to become unicorns (high valued companies) bringing in good returns. On the contrary, there are examples of few startups that failed and eventually closed their businesses due to various issues and challenges.

ISSUES AND CHALLENGES OF STARTUPS

A successful start-up cannot start a business just with passion and an idea. A high level of leadership skills with clear understanding of market, excellent communication skills, maturity to see things in right perspective along with the ability to take calculated risks are required on the part of the entrepreneur(Aggarwal,2017). Lack of awareness, multiple clearances, unorganised market, poor infrastructure in Tier 2 /3 cities, lack of mentoring , stringent exit policies, corruption/red tape, technological risk, regulatory obstacles and lack of reforms keeping pace with the fast evolving market changes are some of the challenges as per Rashmi Guptey, Principal (Legal) of Lightbox India Advisors Private Limited.

Some of the major issues and challenges are discussed below:

Financial Resources

Availability of finance is critical for the startups and is always a problem to get sufficient amounts (Mittal, 2014; Truong, 2016). A number of finance options ranging from family members, friends, loans, grants, angel funding, venture capitalists, crowdfunding etc are available. The requirement starts increasing as the business progresses. Scaling of business requires timely infusion of capital. Proper cash management is critical for the success of the

startups (Skok, 2016;Pandita,2017). A recent report paints a gloomy picture with 85% of new company's reportedly underfunded indicating potential failure (Iwasiuk, 2016).

Revenue Generation

Several startups fail due to poor revenue generation as the business grows. As the operations increase, expenses grow with reduced revenues forcing startups to concentrate on the funding aspect, thus, diluting the focus on the fundamentals of business. Hence, revenue generation is critical, warranting efficient management of burn rate which in common parlance is the rate at which startups spend money in the initial stages. The challenge is not to generate enough capital but also to expand and sustain the growth.

1) Team Members

To find and hire the right kind of talent for the business with skills to match growing customer expectations are one of the biggest challenges (Truong,2016). Apart from founder(s), startups normally start with a team consisting of trusted members with complementary skill sets. Usually, each member is specialized in a specific area of operations. Assembling a good team is the first major requirement, failure to have one sometimes could break the startup (Skok, 2016). According to a survey, 23 percent startups failed because members did not work as a team. Chirag Garg, CEO, HyperDell, feels that **bringing in affordable talent at the right time is a challenge. As per** Nitin Sharma, Principal & Founding member, Lightbox India Advisors Private Limited "Hiring and retaining high quality talent, especially in the areas of product and technology remains a key challenge" (Choudhary,2015)

2) Supporting Infrastructure

There are a number of support mechanisms that play a significant role in the lifecycle of startups which include incubators, science and technology parks, business development centers etc. Lack of access to such support mechanisms increases the risk of failure.

3) Creating Awareness in Markets

Startups fail due to lack of attention to limitations in the markets. The environment for a startup is usually more difficult than for an established firm due to uniqueness of the product. The situation is more difficult for a new product as the startup has to build everything from scratch.

4) Exceed Customer Expectations

The next most important challenge is gauging the market need for the product, existing trends, etc. Innovation plays an important role, since, that the startup has to fine-tune the product offerings to suit the market demands (Skok, 2016). Also, the entrepreneur should have thorough domain knowledge to counter competition with appropriate strategies. Due to new technologies that are emerging, the challenge to provide over and above an earlier innovation is pertinent. Namrata Garg, Director, SendKardo feels that t**he biggest challenge is the need to constantly reinvent yourself and come up with a service to be able to match up customer expectations** and exceed them.

5) Tenacity of Founders

Founders of startups have to be tough when the going gets tough. The journey of starting a venture is fraught with delays, setbacks and problems without adequate solutions. The entrepreneur needs to be persistent, persuasive, and should never give up till he/she achieves desired results. History is replete with startups who gave up the fight when things went wrong. Sometimes the product could be ahead of its time or may require complimentary technology /products for the use by the customers. For example, Apple had to delay introduction of iTunes till the regulations favoured the launch. It is also relevant to quote Steve Jobs who by

commenting "*A lot of times, people don't know what they want until you show it to them"* reiterates the fact those products from startups mostly fall in the "new and untried" category where the success rate is minimal.

6) Regulations

Starting a business requires a number of permissions from government agencies. Although there is a perceptible change, it is still a challenge to register a company. Regulations pertaining to labor laws, intellectual property rights, dispute resolution etc. are rigorous in India which takes about 30 days to comply compared to just 9 days in OECD countries. Also, as per World Bank report, "World Bank Ease of Doing Business", India ranks 142 out of 189 economies (Mittal,2014).

7) Growth Decelerators

Some of the agencies which are part of the startup ecosystem themselves can sometimes become hurdles in the growing stages. As per Sneh Bhavsar, co-founder and CEO, OoWomaniya one of the major issues is **the influence of** incubators, institutes and similar organisations which try to control, manage and be the daddies of the start-ups in the name of helping, mentoring etc (Choudhury,2015).This needs proper coordination among the organizations for mutual benefit.

8) Lack of Mentorship

Milan Hoogan, Vice President -Sales and Marketing at Erfolg Life Sciences feels that l**ack of proper guidance and mentorship** is one of the biggest problems that exist in the Indian startup ecosystem (Choudhury, 2015) . Most of startups have brilliant ideas and/or products, but have little or no industry, business and market experience to get the products to the market. It is a proven example that a brilliant idea works only if executed promptly (Mittal,2014). Lack of adequate mentoring/guidance is the biggest challenge which could bring a potentially good idea to an end.

9) Lack of a Good Branding Strategy

Absence of an effective branding strategy is another issue that prevents startups from flourishing at a faster pace. Hemant Arora, Business Head-Branded Content, Times Network opines that branding demands paramount attention as it gives an identity and occupies a space in the consumer minds(Choudhury,2015) .

10) Replicating Silicon Valley

Koushik Shee, Founder and CEO, Effia **,** feels that Indian startups get influenced by Silicon Valley models which may not succeed in Indian scenario. Lot of tweaking and modifications could be required when transplanted into Indian markets keeping in mind Indian infrastructure in terms of roads, internet, electricity and telecom penetration (Choudhury,2015) .

1. REASONS FOR FAILURE

As regards major reasons for failure of startups, a survey based on analysis of 101 firms showed that 42% failed as the product had no market, 29% firms ran out of cash, 23% did not have the right team,18% closed due to pricing issues, 17% firms had poor product, 14% failed due to poor marketing and 8% had no investor interest(Griffith,2014). These reasons substantiate most of the issues and challenges that have been enumerated above.

GOVERNMENT INITIATIVES

There are numerous government and semi-governmental initiatives to assist startups.

- **Start-Up India**

  This initiative provides three-year tax and compliance breaks intended for cutting government regulations and red tapism.

- *MUDRA Yojna*

Through this scheme, startups get loans from the banks to set up, grow and stabilize their businesses.

- **SETU (Self-Employment and Talent Utilization) Fund**

Government has allotted Rs 1,000 Cr in order to create opportunities for self-employment and new jobs mainly in technology-driven domains.

- **E-Biz Portal**

Government launched e-biz portal that integrates 14 regulatory permissions and licenses at one source to enable faster clearances and improve the ease of doing business in India.

- **Royalty Tax**

Indian government has reduced the royalty tax paid by businesses and startup firms from 25per cent to 10 per cent.

CONCLUSIONS

In this paper we discussed about intership, innovation, incubation and startup with their types with national, international. Our proposed method describes the above parameters for technical institutions in india.The current economic scenario in India is on expansion mode. The Indian government is increasingly showing greater enthusiasm to increase the rate of growth from grass root levels with introduction of liberal policies and initiatives for entrepreneurs like 'Make in India', 'Startup India', MUDRA etc. 'Make in India' is great opportunity for the Indian start-ups. With government going full hog on developing entrepreneurs, it could arrest brain drain and provide an environment to improve availability of local talent for hiring by startup firms. In this paper,as the part of higher education discussion about types of internship to educate the engineers to the suitable working environment in the institutions. Those students has participated interested to learn and digest working environment with academics. while in the internship period the has learn and getting the idea generation to produce the product out come. Small contributions from a number of entrepreneurs would have cascading effect on the economy and employment generation which would complement medium and large industries efforts catapulting India into a fast growing economy. The startup arena has lot of challenges ranging from finance to human resources and from launch to sustaining the growth with tenacity. Being a country with large population, the plethora of opportunities available are many for startups offering products and services ranging from food, retail, and hygiene to solar and IT applications for day to day problems which could be delivered at affordable prices. It is not out of place to mention that some of these startups would become unicorns and may become world renowned businesses by expanding into other developing and underdeveloped countries.

REFERENCES

1. Brown, S., Bucklow, C., Clark, P., 2002. Professionalizing teaching: enhancing the status of teaching, improving the experience of learning and supporting innovation in higher education. J. Geogr. High. Educ. 26 (2), 159–168.

2. Amason, A. 1996. Distinguishing the effect of functional and dysfunctional conflict on strategic decision making: Resolving a Paradox for top management teams. Academy of Management Journal, 39 pp. 123-148.
3. Bichsel, J. 2012. Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations. [e-book] EDUCAUSE Center for Applied Research. Available at: http://www.educause.edu/eca.
4. Bohle Carbonell, K. et al. 2012. Unleashing the creative potential of faculty to create blended learning. The Internet and Higher Education.
5. Brennan, J. et al. 2013. Things we know and don't know about the Wider Benefits of Higher Education: A Review of the Recent Literature. BIS Research Paper Number 133.
6. Venkatesh, U., & Dutta, K. (2007). Balanced scorecard in managing higher education institutions: An India perspective. International Journal of Educational Management, 21(1), 54–67.
7. Vishnoi, A. (2013). Innovation Universities Bill Stuck, University Grants Commission Looks at Giving Funds to Existing Universities for Innovation, The Indian Express. Retrieved from http://www.indianexpress.com/news/innovation-universities-bill-stuck-universi/1068853/.
8. Spady, William (1994). *Outcome-Based Education: Critical Issues and Answers*(PDF). Arlington Virginia: American Association of School Administrators. ISBN 0876521839. Retrieved 31 October 2014.
9. Mohayidin, Mohd Ghazali (2008). "Implementation of Outcome-Based Education in Universiti Putra Malaysia: A Focus on Students' Learning Outcomes". *International Education Studies*. 1 (4). Retrieved 23 October 2014.
10. Kaliannan, Maniam; Chandran, Suseela Devi (2012). "Empowering Students through Outcome-Based Education (OBE)". Research in Education. 87 (1): 50–63.
11. Desai, S., & Kurkarni, V. (2008). Changing educational Inequalities in India in the context of affirmative action. Demography, 45(2), 245–270.
12. Dube, S. (1988). Higher education and social change. In A. Singh & G. Sharma (Eds.), Higher education in India: The social context (pp. 120–176). New Delhi: Konark.
13. Bloom, B. S. (1994). "Reflections on the development and use of the taxonomy". In Rehage, Kenneth J.; Anderson, Lorin W., Sosniak, Lauren A. (eds.). *Bloom's taxonomy: A forty-year retrospective*. Yearbook of the National Society for the Study of Education. **93**. Chicago: National Society for the Study of Education. ISSN 1744-7984.
14. http://www.kurwongbss.qld.edu.au/thinking/Bloom/blooms.htm.
15. Anderson, Lorin W. & Krathwohl, David R. (2001). A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy. New York. Longman Publishing.
16. Clark, Donald R. (1999). "Bloom's Taxonomy of Learning Domains". Retrieved 28 Jan 2014.

17. Hoy, Anita Woolfolk (2007). Educational psychology (10th ed.). Boston: Pearson/Allyn and Bacon. pp. 530–531, 545. ISBN 0205459463. OCLC 68694368.
18. Armstrong, Patricia. "Bloom's Taxonomy". Vanderbilt University Center for Teaching. Vanderbilt University. Retrieved 29 June 2016.
19. Harrow, Anita J. (1972). A taxonomy of the psychomotor domain: A guide for developing behavioral objectives. New York: David McKay Company.
20. Dave, R. H. (1975). Armstrong, R. J. (ed.). Developing and writing behavioral objectives. Tucson: Educational Innovators Press.
21. M. Yang, M. You and F. Chen "Competencies and qualifications for industrial design jobs: implications for design practice, education, and student career guidance" Design Studies, Vol. 26, Elsevier Ltd, UK, (2005), 155-189.
22. W. Lewis, and E. Bonollo, "An analysis of professional skills in design: implications for education and research", Design Studies, Vol. 23, Elsevier Ltd, UK, (2002), 385-406.
23. A. Salama, "A Voice for an alternative architectural education: integrating "What" and "How" knowledge" ARCHITTIMES, Karachi, Pakistan, September (2003).
24. Agarwal, P. (2009). Indian higher education: Envision the future. New Delhi: SAGE.
25. Brown, P., Lauder, H., & Ashton, D. (2010). The global auction—The broken promises of education, jobs, and incomes. Oxford: Oxford University Press.
26. Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. London: SAGE.
27. Government of India. (1958). Scientific policy resolution. In Government of India. Department of
28. Science and Technology (Ed.). New Delhi: Ministry of Science and Technology.
29. Government of India. (1983). Technology policy statement. In Department of Science & Technology (Ed.). New Delhi: Ministry of Science and Technology.
30. Government of India. (2003). Science and technology policy. In Department of Science & Technology (Ed.). New Delhi: Ministry of Science and Technology.
31. Government of India. (2012). Twelfth five year plan (2012–2017) social sectors. In Planning Commission (Ed.), (Vol. 3, p. 274). New Delhi: SAGE.
32. Government of India. (2013). Science, technology and Innovation policy. In Minister of Science & Technology (Ed.), (p. 22). New Delhi: Ministry of Science and Technology.
33. Bound, K., & Thornton, I. W. (2012). Our frugal future: Lessons from India's innovation system. London: NESTA.
34. Kaul, S. (2006). Higher Education in India: Seizing the Opportunity Working Paper Series (Vol. 179). New Delhi: Indian Council for Research on International Economic Relations.

Low Overhead Warning Flip-Flop Based on Charge Sharing for Timing Slack Monitoring

POCHANA ASHOK KUMAR , MTECH(ES&VLSI) , MRCE

***Abstract*— Timing error predictors have a strong potential to reduce the worst case timing margins by monitoring timing slack of a design. However, these timing error predictors incur substantial amount of silicon area and power which limit the overall benefits in the system level. This paper presents a low overhead warning flip-flop (FF), which predicts setup time violations. It consists of a delay buffer and a warning generator along with a conventional master-slave FF. Low overhead FF can be designed by exploiting the concept of charge sharing to implement the warning generator. As the warning generator requires only seven transistors to predict the timing violation, the proposed warning FF occupies 30% less area and consumes27% less power compared to the state-of-the-art timing error predictors. A test chip is fabricated using the proposed FF in a 130-nm CMOS technology to verify the functionality of the proposed warning FF in dynamic voltage and frequency scaling applications. Measurement results from the test chip show that a performance improvement of 44% can be achieved at a supply voltage of 0.9 V by employing the proposed technique compared to the worst case design. For a typical chip, the power consumption can be reduced by 36% compared to the worst case design.**

***Index Terms*— Charge sharing, dynamic frequency scaling (DFS), dynamic voltage scaling (DVS), setup time violation, timing error predictors, timing margins.**

I. Introduction

PROCESS, voltage, and temperature (PVT) variations in scaled technology nodes cause significant performance uncertainty in the digital designs. Timing or supply voltage guard bands are added to maximum operating frequency (MOF) or minimum supply voltage to cope with PVT variations. However, these guard bands severely limit the performance and/or increase the power consumption of a design in the typical or best conditions. Moreover, transistor aging degrades the performance of a design with time. Hence, guard bands need to be added considering the life time of the design. As a result, traditional worst case design methodology is not suitable to implement energy-efficient designs as large guard bands are required in nanometer technology nodes. This motivates to implement design methodologies which can reduce the guard bands.

Traditionally, process monitors have been proposed to monitor the manufacturing process condition of a chip. In these techniques, body biasing is used to change the thresh- old voltage of the transistors based on process conditions. However, these techniques can address only global process variations but not local process and dynamic variations. Critical path replica circuits [1]–[4] have been explored to track the delay of the critical paths of a design. In this approach, a replica circuit which is strongly correlated with the actual critical path of the design is monitored to observe delay variations. This method can address global varia- tions but not the local variations due to mismatch in the delay of the actual critical path and replica path. Moreover, the activation of critical path depends on the input data pattern.

On the other hand, *in situ* timing error monitoring techniques [5]–[8] can address both local and global variations by monitoring the timing slack. These techniques directly monitor the output of combinational logic using a specialized flip-flop (FF). An error signal is flagged in the case of timing violation. Hence, the supply voltage or frequency of design can be altered by monitoring the error signal. These techniques are mainly classified into two categories: error detectors [5], [6] and error predictors [8], [9]. Error detectors such as Razor I [6] and Razor II [7] detect the timing errors after their occurrence and correct the timing violations using architectural replay mechanism [7]. However, error detectors introduce a significant minimum path delay constraint which causes large area overhead because of buffer insertion. Moreover, architectural replay mechanism employed to correct timing violations is available in high-performance processors but not in application-specific integrated circuits (ASICs). Bubble Razor [10] does not incur area overhead for short path padding. However, this technique uses latches in the pipelines instead of FFs.

Error predictors [11], [12] flag a warning signal before the occurrence of timing violations by monitoring the delayed data. As the output of FF is always correct, these techniques do not incur correction overhead. Error predictors are suitable to implement ASICs as they do not require a correction mechanism. However, these techniques can monitor only gradual change in the delay of the critical paths. Canary FF [8], [9], [11] falls into this category. Canary FF employs a double-sampling architecture to predict the timing violations. Canary FF incurs large area and power overhead because of the shadow FF and delay buffers. Moreover, output of the shadow FF may enter into metastable state.

In [13], an aging sensor has been proposed, which uses delayed clock to create a guard band interval before the rising edge of the clock. As the guard band interval is created using the delayed edge of the previous clock cycle, the design is a function of operating frequency. An alternative version is also proposed in [13], which uses double-sampling architecture similar to canary FF. A timing monitoring circuit presented in [14] uses a sensor to predict the timing violations and a warning window generator to create a detection window. In this approach, the detection window is generated by using a transition detector in the clock tree which makes the clock tree implementation difficult.

A warning detection sequential in [12] and [15] observesthe delayed data transition during a warning interval to predict timing violations. This FF is not area and power efficient as a large number of delay buffers are required to delay the input data and to implement the edge detector. In [16], a sensor has been proposed which monitors the delayed master output to predict the timing violations. It requires a large area and power as it uses double-sampling architecture. A pre-error FF in [17] monitors data transition during the negative half cycle of the clock cycle. As warning margin is determined by negative half cycle of the clock, the design requires a clock of more than 50% duty cycle for better performance of the design.

The pulse-based error predictor in [18] monitors delayed master latch output in the high phase of the clock. However, edge detector used in this design requires more area and consumes more power. *In situ* monitors proposed in [19]–[21] observe master latch output and employ double-sampling architecture similar to the canary FF [8]. A timing error predictor in [22] employs double-sampling architecture similar to the canary FF except that a tunable delay buffer is used to delay the data signal. Double-sampling architectures require significant area and power overhead. A current-based timing error detector is proposed in [23]. A nine transistor transition detector [24] is designed to operate at supply voltage range 0.44–1.1 V for low-power applications. Critical path identification becomes difficult with increased variations in nanometer technology nodes [25]. Because of this, more number of critical paths have to be monitored using the timing error predictors to avoid functional failure. Hence, this paper presents a low area and power overhead timing error predictor for timing slack monitoring. In addition, the proposed warning FF can be used as an aging sensor similar to the work reported in [13] and [26].

The remainder of this paper is organized as follows.

The operation of the proposed warning FF is discussed in Section II. Chip implementation details are described in Section III. Measurement results are presented in Section IV. Section V discusses the comparison of timing error predictors and simulation results. Finally, this paper is concluded in Section VI.

II. Proposed Charge Sharing-Based Warning Flip-Flop

Warning FF is used to predict the setup time violation. Warning FF flags a warning signal, if data transitions at the input of FF happen during a timing window before the rising edge of the clock. Traditionally, either the input data of FF [11], [15] or master latch output [18], [20] is monitored to predict the setup time violations. In both approaches, the concept of delayed data is employed to predict the timing violations. Moreover, these approaches use either double- sampling technique or transition detection technique to flag warning signal. Based on these, warning FFs can be classified into four categories:

1) samples at input data *D* of FF and uses double-sampling technique [11];
2) samples master latch output *Mout* and uses double-sampling technique [20];
3) samples at input data *D* of FF and uses transition detection technique [15];
4) samples master latch output *Mout* and uses transition detection technique [18].

If the input data of the FF is monitored to flag warning signal, the delay of buffer should account for both setup time of FF and warning margin. However, if the master latch output is monitored to flag warning signal, the delay of buffer should account for warning margin only as the master latch delay already accounts for the setup time of FF.

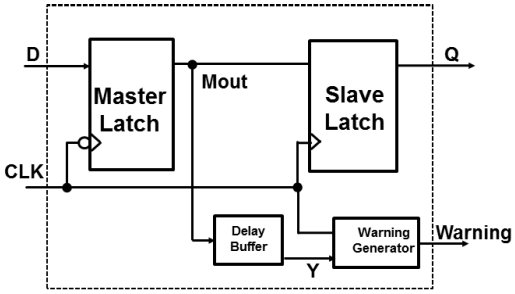


Fig. 1. Proposed warning FF.

The proposed warning FF is based on monitoring the delayed data at the output of master latch. It consists of a delay buffer and a warning generator along with a conventional master-slave FF to predict the setup time violations. The delay of the buffer is the warning margin of the proposed FF. If a data transition happens during a timing interval equal to the warning margin before the setup time window of the master latch, the proposed FF flags the warning signal. The warning signal is an active low signal. In the case of early data arrival at the input of FF, the warning generator output is logic high. However, in the case of late data arrival, the warning generator output is logic low. The schematic of the proposed warning FF is shown in Fig. 1.

A. Operation of the Proposed Flip-Flop

A data transition at the input of the FF can happen in either of the four timing windows shown in Fig. 2. A data transition during the window 1 is treated as an early data arrival. In this case, the available timing slack is very high. A data transition during window 2 is treated as late data arrival. In this case, the available timing slack is less. A data transition in this window means that a timing violation may occur if the delay of the critical path is further increased due to PVT variations. A data transition in timing window 3 leads to setup violation,

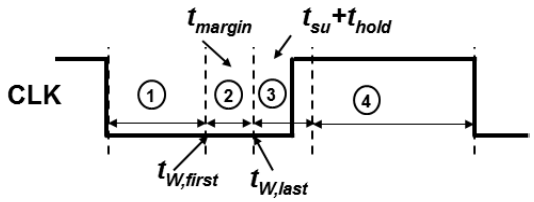


Fig. 2. Timing windows for possible data transition.

TABLE I
OPERATION OF THE PROPOSED WARNING FF

Timing Window	Data Transition	Before the rising edge of clock			After the rising edge of clock			Q	Warning
		D	Mout	Y	D	Mout	Y		
1	Rise	1	1	1	X*	1	1	1	1
	Fall	0	0	0	X	0	0	0	1
2	Rise	1	1	0	X	1	1	1	0
	Fall	0	0	1	X	0	0	0	0
3	Rise or Fall	X	PS**	PS	X	PS/D	PS/D	PS/D	1/0
4	Rise or Fall	X	PS	PS	X	PS	PS	PS	1

* Don't care ** Previous State

and the output of master latch may enter into metastable state. So, the maximum delay constraint of the design should satisfy (1), so that the data transitions under PVT variations can happen before the timing window 3. The maximum delay constraint is given by

$$T_{\max} \leq T_{\mathrm{CLK}} - t_{\mathrm{ClktoQ}} - t_{\mathrm{su}} - t_{\mathrm{margin}} \quad (1)$$

where $T_{\max}$ is the maximum combinational path delay, T_{CLK} is the clock period, t_{ClktoQ} is the previous stage FF clock to Q delay, t_{su} is setup time, and t_{margin} is warning margin. A data transition during the timing window 4 does not appear at the output of master latch as the master latch is opaque in this window.

The operation of the warning FF is given in Table I. If a data transition occurs during the timing window 1, the master latch output (*Mout*) and delayed master latch output (*Y*) (refer to Fig. 1) are the same in both low and high phases of the clock. However, if a data transition happens in the timing window 2, the delayed master output (*Y*) is different from *Mout* in low and high phases of the clock. In this case, the warning signal becomes low. However, in both the cases, the FF samples the correct value. When a data transition happens during the timing window 3, the master latch output (*Mout*) may enter into metastable state and it can resolve to either logic "0" or logic "1". Hence, the delayed master output (*Y*) may store the present data (*D*) or the previous state of FF. The conceptual timing diagram of the proposed warning FF is shown in Fig. 3.

In clock cycle I, the data transition happens early, before the warning margin. The master latch samples the correct value. The delayed master latch output (*Y*) makes a transition during the negative half cycle of the clock represented as *A*1. In this case, delayed master output (*Y*) is the same in both the low and high phases of the clock. Hence, the warning generator output is logic high signal. The delay between the input of FF (*D*) and the master latch output (*Mout*) is the setup time of master latch. In clock cycle II, data transition

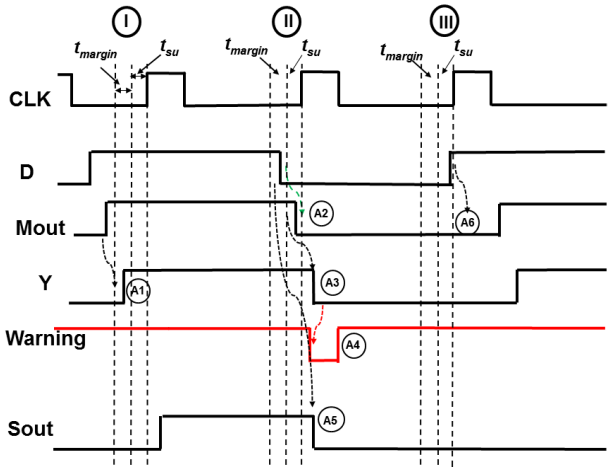


Fig. 3. Conceptual timing diagram of the proposed warning FF.

happens during the warning margin interval, the master latch samples the valid data represented as *A*2. However, the delayed master latch output makes transition during the high phase of the clock represented as *A*3. In this case, the delayed master output is different in high and low phases of the clock. So, the warning generator output becomes low represented as *A*4. Even the warning signal is flagged, the output of FF is always correct represented as *A*5. So, there is no correction overhead with this approach. In clock cycle III, the data transition happens during the setup time window of the master latch. This data transition does not appear at the output of master latch represented as *A*6. Hence, the proposed FF does not detect data transitions happened during this window. Similarly, data transitions during the high phase of clock will not appear at the master latch output as it is opaque during the high phase of the clock. Hence, the proposed FF does not suffer from short path problem. The minimum delay constraint of the design is given by

$$T_{\min} \geq t_{\mathrm{hold}} - t_{\mathrm{ClktoQ}} \quad (2)$$

where $T_{\min}$ is the minimum combinational path delay, t_{ClktoQ} is the minimum clock-to-Q delay of the previous stage FF, and t_{hold} is the hold time of FF.

The detectable slack of the proposed warning FF is defined as the difference between the point-of-first-warning (POFW) ($t_{W,\mathrm{first}}$) and the point-of-last-warning ($t_{W,\mathrm{last}}$) (refer to Fig. 2). Ideally, the detectable slack is the delay of buffer, which is used to delay the master latch output. Hence, the detectable slack is given by

$$t_{d,\mathrm{slack}} = t_{d,\mathrm{buffer}} \quad (3)$$

where $t_{d,\mathrm{slack}}$ is the detectable slack and $t_{d,\mathrm{buffer}}$ is the delay of the buffer.

B. Warning Generator

The warning generator is designed using the concept of charge sharing. It requires only seven transistors to detect a transition during the high phase of the clock. The schematic of the warning generator is shown in Fig. 4. The delayed master output drives two cascaded-clocked inverters. During the low

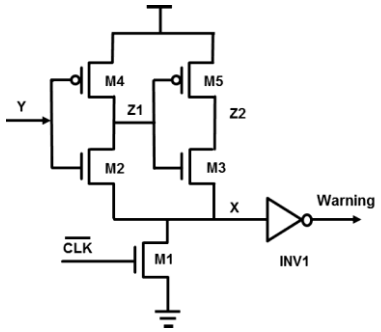


Fig. 4. Schematic of the warning generator.

TABLE II
OPERATION OF WARNING GENERATOR

Timing Window	Data Transition	CLK=0				CLK=1				Warning
		Node Y	Node Z1	Node Z2	Node X	Node Y	Node Z1	Node Z2	Node X	
2	Rise	1	0	1	0	1	0	1	0	1
	Fall	0	1	0	0	0	1	0	0	1
3	Rise	0	1	0	0	1	Discharged	Charged	Charged	0
	Fall	1	0	1	0	0	Charged	Discharged	Charged	0
4	Rise	0	1	0	0	0	1	0	0	1
	Fall	1	0	1	0	1	0	1	0	1
	Rise	0	1	0	0	0	1	0	0	1
	Fall	1	0	1	0	1	0	1	0	1

phase of the clock, these inverters behave like conventional inverters and the intermediate node *X* is discharged to logic zero. However, during the high phase of the clock, node X becomes floating as the transistor M1 is OFF. The warning signal is at logic high state. Whenever the input of warning generator is different in low and high phases of the clock, the intermediate node *X* charges from the nodes *Z*1 and *Z*2 for rise and fall data transitions, respectively. However, the charged voltage at intermediate node *X* is less than supply voltage for rise and fall transition at node *Y*. The inverter INV1 should treat the charged voltage at node *X* as logic high-input voltage. For this, the switching threshold of inverter should be less than charged voltage at node *X*. The inverter INV1 is skewed to change the switching threshold. The switching threshold of the inverter is less than half of the supply voltage.

The operation of the warning generator is given in Table II. The delayed master output (*Y*) depends on timing window (refer to Fig. 2) in which input data (*D*) transition happens. An input data (*D*) transition may lead to: 1) a transition at *Y* during the low phase of the clock; 2) a transition at *Y* during the high phase of the clock; and 3) no transition until the next low phase of the clock. A transition at *Y* during the low phase of the clock occurs, if the data (*D*) transition happens in timing window 1. A transition at *Y* during the high phase of the clock occurs, if the data (*D*) transition happens in timing window 2. For a data transition in timing windows 3 and 4, no data transition happens at *Y* until the next low phase of the clock. The detailed operation of the warning generator is discussed in Sections II-B.1–II-B.3.

1) Data (Y) Transition During Clock Low Phase: If a signal transition at *Y* occurs during the low phase of the clock, the state of the node *Y* is the same in both low and high phases of the clock. So, the intermediate nodes *X*, *Z*1, and *Z*2 do not get affected after the rising edge of the clock. When CLK = 0, the intermediate node *X* is discharged. Hence, the warning signal is high.

2) Rise Transition (Y) During Clock High Phase: For a rise transition at the input of the warning generator (*Y*) in the high phase of clock, the warning generator input is initially logic zero. A logic zero input during the low phase of clock, charges the intermediate node *Z*1 and discharges the intermediate node *Z*2. The initial states of nodes *X*, *Z*1, and *Z*2 are shown in Fig. 5(a). The transistor M1 is ON, during the low phase of the clock. So, the node *X* is discharged to logic zero. In addition, the nodes *Z*1 and *Z*2 are logic high and low, respectively. After a rise transition at the input of warning generator in the positive half cycle of clock, the states of nodes *X*, *Z*1, and *Z*2 are shown in Fig. 5(b). In this case, the transistor M1 is *off*. So, the intermediate node *X* is charged through the transistor M2. Because of charge sharing between nodes *X* and *Z*1, the voltage at node *X* starts increasing toward logic high. If this voltage is more than the switching threshold of the inverter, the warning signal becomes low. In this way, a rise transition at delayed master output is detected.

3) Fall Transition (Y) During Clock High Phase: For a fall transition at *Y* in the high phase of the clock, the delayed master output should be logic high in the negative half cycle of the clock. The initial states of nodes *X*, *Z*1, and *Z*2 are shown in Fig. 5(c). The transistor M1 is ON during the low phase of the clock, so the warning signal is high during this interval. The intermediate nodes *Z*1 and *Z*2 are logic low and high, respectively. After a fall transition in the high phase of the clock, the node *X* is charged from node *Z*2 through transistor M3. This charge sharing leads to a rise transition at node *X*. The output of warning generator becomes low. The states of nodes *X*, *Z*1, and *Z*2 in the high phase of the clock are shown in Fig. 5(d).

III. CHIP IMPLEMENTATION DETAILS

A two-stage pipelined design is implemented to verify the functionality of the proposed warning FF. The internal blocks of the test chip is shown in Fig. 6. The implemented blocks are divided into three sections: 1) input section; 2) pipelined design; and 3) output section. The input section consists of a serial-in parallel-out (SIPO) shift register and a toggle circuit. The pipelined design section consists of two-stage pipelined circuit with the proposed warning FF and an AND tree to group the warning signals. The output section consists of a parallel-in serial-out (PISO) shift register.

A 13-bit SIPO shift register is used in the implementation of the design. Out of 13 bits, 8 bits for input data of the pipelined design, 1 bit for enable signal of the toggle circuit, 2 bits for selecting the critical path delay, and 2 bits for selecting the warning margin of the warning FF are assigned. To check the functionality of the proposed warning FF on chip, the critical path needs to be activated in each clock cycle, which can be done by changing the inputs of the pipelined design using a toggle circuit. The data input to the design is given from an SIPO shift register through the toggle circuit. Toggle circuit has an enable signal. If the enable signal is high, the inputs of the design will change in every clock cycle which is required to evaluate the warning FF in the pipelined design. In this case, the input data pattern stored in the SIPO register can be inverted in each clock cycle using the toggle circuit and

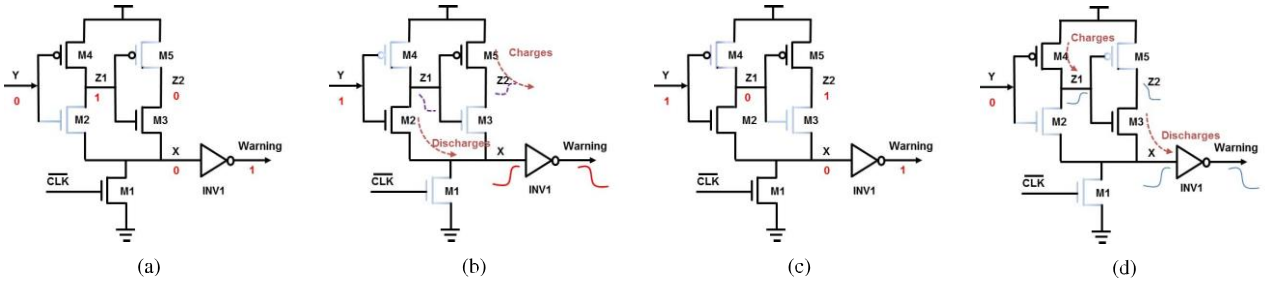


Fig. 5. Operation of warning generator for data transitions in timing window 2. (a) Before rise transition. (b) After rise transition. (c) Before fall transition. (d) After fall transition.

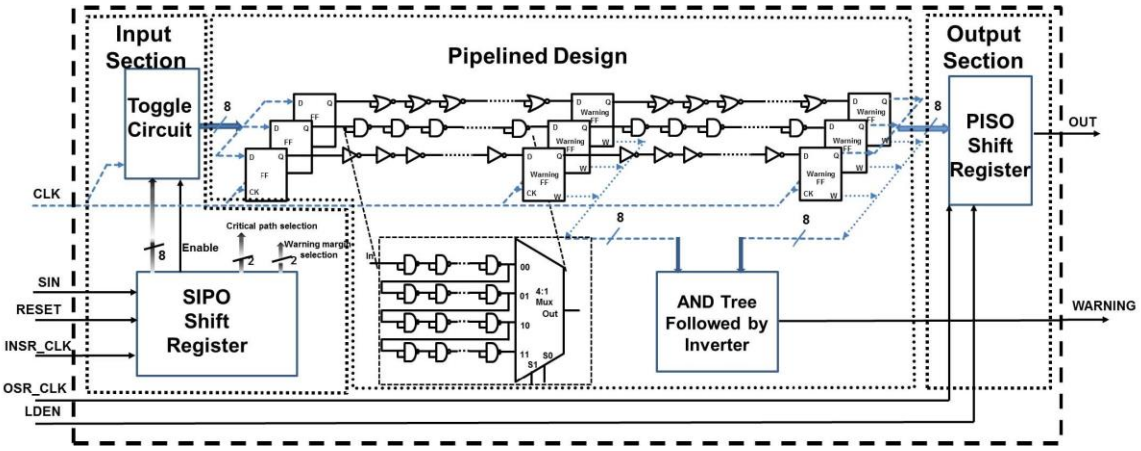


Fig. 6. Details of chip implementation.

is provided to the pipelined design. As all the paths in the pipelined design are made critical at the design time, all the paths have 100% activity during run time.

Each stage of the pipelined design has eight paths which are designed using inverter, NAND and NOR chains. The proposed warning FF is inserted at the end points of all the critical paths. The length of the critical path can be modified inside the chip using a configurable delay chains with a 4:1 multiplexer in the critical path. Hence, this scheme allows operating the same design at different clock frequencies. The warning signal from all the warning FFs are grouped into a single warning signal using an AND tree. An inverter converts the AND-tree output to an active high warning signal WARNING as the warning signal in the FF level is active low. The delay of the AND tree should be less than the minimum pulsewidth of the warning signal of the proposed FF and is given by

$$t_{d,\text{AND-tree}} = T_{\text{PW,min}} \quad (4)$$

where $t_{d,\text{AND-tree}}$ is the delay of AND tree and $T_{\text{PW,min}}$ is the minimum pulsewidth of the warning signal of the proposed FF.

An 8-bit PISO shift register is used to load the outputs of the pipelined design. The PISO shift register has two external inputs: one is the clock (OSR_CLK) to the shift register and the other is a control signal (*LDEN*) to enable the shift register to load from the pipelined design and to shift the data of the shift register. If the LDEN is high, the outputs of the pipelined design are loaded into PISO shift register in every clock cycle.

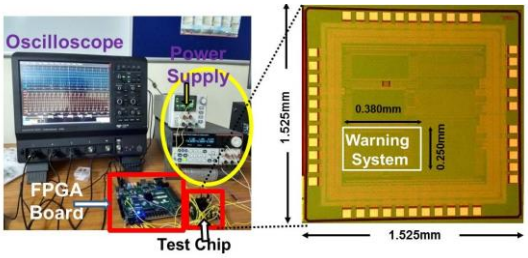


Fig. 7. Experimental setup.

If the LDEN is low, the stored data of PISO shift register is serially shifted out in every clock cycle.

IV. Measurement Results

The experimental setup to measure the test chip is shown in Fig. 7 which shows a test chip implemented in 130-nm CMOS technology, mounted on QFN-48 socket, a field-programmable gate array (FPGA) board and an oscil- loscope. The SIPO shift register inside the chip is programed through the FPGA Board (ZedBoard) to provide the input data. The serial output of the PISO register is monitored continuously. In the case of warning, all the outputs of the design are verified by shifting the PISO register.

The frequency of baseline or worst case design with 10% voltage drop, a temperature of 85 °C, and 2-sigma process variations is obtained by using the method of [22].

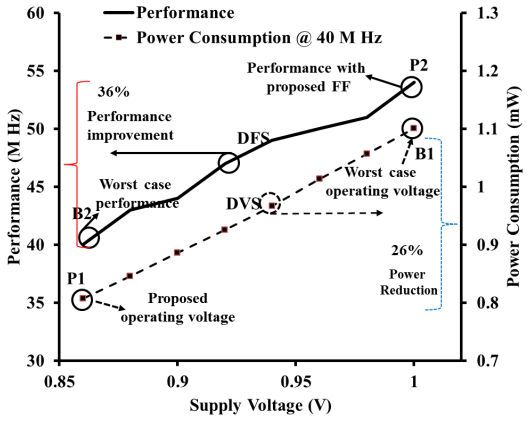


Fig. 8. Performance and power consumption of the design.

The frequency of the baseline design is 40 MHz for a supply voltage of 1 V. The conventional operating voltage and baseline frequency of the design are represented as *B*1 and *B*2, respectively, in Fig. 8. However, by employing the proposed FF in the pipelined design with dynamic voltage scaling (DVS), the pipelined design can operate at a reduced voltage represented as *P*1 for the same performance which results in 26% savings in the power consumption. For dynamic frequency scaling (DFS), the supply voltage is kept constant and frequency of the design is increased. The pipelined design with the proposed FF and DFS can operate at an increased frequency represented as *P*2 which leads to 36% performance improvement at a supply voltage of 1 V. This shows the advantage of using the proposed FF in reducing the timing or voltage guard bands.

A. Impact of Supply Voltage Scaling

Conventionally, the design will operate at a supply voltage of 1.2 V considering worst case margins. However, the pro- posed warning FF can detect the actual operating condition which in turn can reduce the worst case guard bands. Initially, the designs are operated at a supply voltage of 1.2 V with an operating frequency of 50 MHz. The supply voltage of the design is reduced until the warning signal is asserted. The supply voltage for which warning signal becomes high is defined as POFW voltage. If the supply voltage of the design is further reduced, the warning signal is still high and a functional error may occur in the design. A system failure is verified by monitoring the outputs of the design using PISO shift register. The voltage for which a functional failure occurs is known as point-of-first-failure (POFF) voltage. The differ- ence between the POFF and POFW voltages is the available margin for warning. This warning margin is a function of number of delay buffers used to delay the master latch output of the proposed warning FF.

The POFW voltage is measured for 26 chips by reducingthe supply voltage manually in steps of 1 mV until the warning signal becomes high. As the implemented test chip does not include an adaptive controller, the supply voltage is reduced manually while monitoring the warning signal.

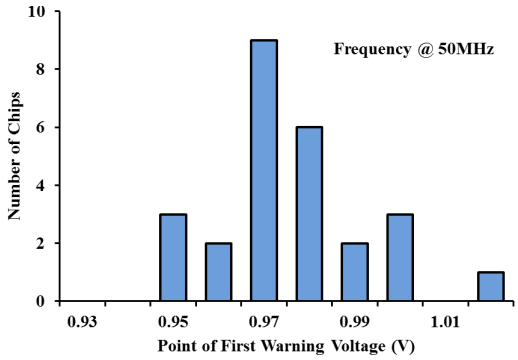


Fig. 9. POFW voltage of measured chips.

However, an external controller used in [15] or on-chip controller proposed in [22] can be employed to change the voltage or frequency of the design. Fig. 9 shows histogram of the POFW voltage for 26 measured chips operating at 50 MHz and at a temperature of 25 °C. It is evident from the POFW voltage of measured chips that the design can operate at reduced supply voltage for a given performance compared to a worst case design. For a typical chip among the measured chips, the supply voltage can be reduced to 0.97 from 1.2 V. However, the minimum voltage required for the worst chip among the measured chips is 1.015 V. Among 26 chips,14 chips can operate at a supply voltage less than 0.97 V for the same performance. Moreover, the best chip among the measured chips requires a supply voltage of 0.94 V, which is 75 mV less than that of the worst chip. This reduction in sup- ply voltage results in lower power consumption, and it shows the significance of warning FF for low-power applications.

To measure the dynamic power consumption of the design (P_{Design}), the following steps are used. First, the power consumption of the design along with input-output (IO) cells (P_{D_IO}) is measured. Then, the power consumption of only IO cells (P_{IO}) is measured by not applying the clock to design. However, three IO cells (clock, output, and warning signals of design) are active, only when the design is working. So, the power consumed by IO cells (P_{IO}) does not include the dynamic power of these three IO cells. So SPICE simulations are carried out to find the dynamic power of these IO cells ($P_{\text{SIM\_IO}}$). To increase the accuracy in the measurement of power con- sumption, the dynamic power of three IO cells ($P_{\text{SIM\_IO}}$) is subtracted. The dynamic power consumption of the design is given by

$$P_{\text{Design}} = P_{D\_IO} - P_{\text{IO}} - P_{\text{SIM\_IO}}. \quad (5)$$

The abovementioned method is used to measure the power consumption in this paper as the same supply voltage is connected for IO cells and the proposed design. The implemented design is having a dedicated clock signal. The histogram of dynamic power consumption of the design for 26 measured chips is shown in Fig. 10. The chips are operated at the POFW to measure the power consumption of the design. The maximum power consumed by the design is 1.35 mW, whereas the minimum power consumed by the design is 1.20 mW.

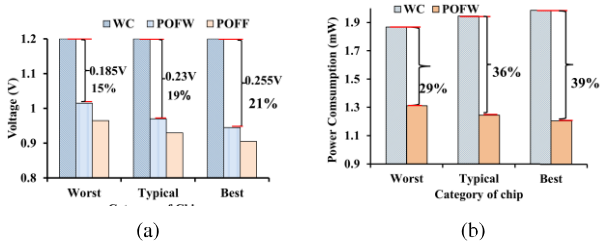


Fig. 11. Comparison of worst, typical, and best chips among the 26 measured chips. (a) POFW voltage. (b) Power consumption.

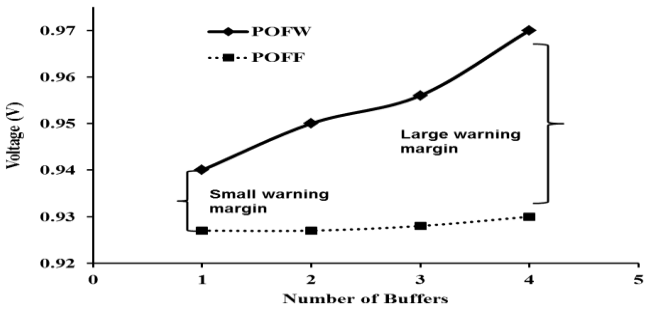
The power consumption of a typical chip among the measured chips is 1.26 mW. Among 26 measured chips, 15 chips can work with a power consumption less than 1.26 mW.

The POFW and POFF voltages for three different categories of chips are shown in Fig. 11(a). The POFW voltage of the worst and typical chips, out of measured chips is 185 and230 mV less than the nominal supply of 1.2 V, respectively. The power consumption of the proposed design with DVS is measured by operating the design at a POFW voltage. For a typical chip, the power consumption is reduced by 36%, whereas for the best chip, the power consumption is reduced by 39% using the proposed warning FF with DVS. The power consumption of three different categories of chips is shown in Fig. 11(b).

B. Impact of Delay Buffers on POFW Voltage

In the proposed approach, a tunable delay buffer used in [15] is employed to change the warning margin of the proposed FF. This tunable delay buffer can be used to provide different warning margins for varying process conditions. A process monitor can be used to detect global process conditions. The warning margin of the proposed FF is varied using the select lines of the multiplexer. In this case, a 4:1 multiplexer is employed for selecting the warning margin.

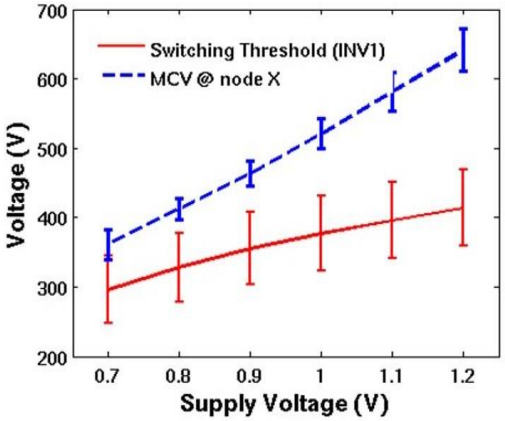
For a select input of 00, the critical path delay is minimum and POFW occurs at a voltage of 0.94 V for a typical chip. In this case, maximum power savings can be achieved. However, the design can tolerate less delay variations as the warning margin is small. For a select input of 11, the warning margin is maximum and the POFW occurs at a voltage of 0.97 V for a typical chip. However, in this case, the design can tolerate large delay variations as the warning margin is maximum. In all the cases, the POFF voltage is the same, as it is determined by the setup time of the FF. Fig. 12 shows the POFW and POFF voltages for a typical chip as a function of delay buffers.



C. Impact of Frequency Scaling

The frequency of baseline or worst case design at a supply voltage can be obtained by using the method of [22]. First, the MOF of each chip is measured at room temperature with 10% drop in the supply voltage. Second, Gaussian fitting of measured MOF of all chips is used to find the mean (μ) and standard deviation (σ). Then,2-sigma frequency (2σ) and 5% of the mean value are reduced from the mean value to account for process variations and temperature margin of 60 °C, respectively. As the design is measured at room temperature, the timing margin to account for the temperature of 85 °C is found by SPICE simulations. It is found from the simulation results that the performance of the design at a temperature of 85 °C is degraded by 6.4% and 5% at a supply voltage of 1.05 and 0.9 V, respectively, compared to performance at room temperature (25 °C). For an operating voltage of 1 V, the MOF of the measured chips at room temperature with 10% voltage drop is shown in Fig. 13. The mean and standard deviation are found to be 45 and1.7 MHz, respectively. The baseline frequency is 40 MHz atan operating voltage of 1 V. Similarly, the baseline frequencies for supply voltage of 0.9, 0.95, and 1.05 V are 31, 35, and44 MHz, respectively.

The MOF of three categories of the chips (worst, typical, and best) is shown in Fig. 14. Out of 26 measured chips, the chip which flags a warning signal for the lowest frequency at a fixed voltage is treated as the worst chip. Similarly, the chip which flags a warning signal for the highest frequency at a fixed voltage is treated as the best chip. The best chip can perform 23% better than a worst chip at a supply voltage of 0.9 V. The same chip can perform 14% better than a worst chip at a supply voltage of 1.05 V. This shows the advantage of using warning FF at lower supply voltage. The proposed warning FF can improve the performance of the design by 44% and 31% compared to baseline design at a supply voltage of 0.9 and 1.05 V, respectively, under typical operating conditions.



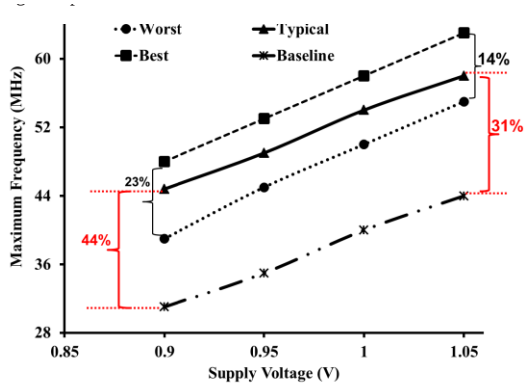


Fig. 14. Performance improvement of worst, typical, and best chips at room temperature.

V. Simulation Results

Section IV discussed the advantages of using the proposed warning FF. The implementation details of warning FF, such as the maximum charged voltage (MCV) at the intermediate node, and sizing of the transistors in the warning generator are discussed in this section. Moreover, the FF-level comparison is discussed to quantify area and power savings with respect to the warning FFs available in the literature. The proposed FF is implemented in industrial 130-nm CMOS technology. Postlayout simulations on parasitic extracted netlist have been carried out using HSPICE.

A. Warning Generator

The warning generator (refer to Fig. 4) is designed using the concept of charge sharing. The MCV at node *X* of the warning generator and switching threshold of the skewed inverter are critical factors for the proper functionality of warning FF. To verify the MCV at node *X* for data transitions at the input of warning generator, 10 000 Monte Carlo simulations have been performed with 3-sigma process variation for a supply voltage range 0.7–1.2 V at a temperature of 25 °C. The minimum value of MCV is used to determine the switching threshold of the skewed inverter INV1 (refer to Fig. 4). The MCV at node *X* and switching threshold of the skewed inverter (INV1) are shown in Fig. 15.

Postlayout simulations on parasitic extracted netlist have been carried out to verify the operation of the proposed

Fig. 16. Timing diagram of the warning generator.

warning generator. Fig. 16 shows the timing diagram of CLKB (inverted clock), input of warning generator (*Y*), warning signal, intermediate nodes *X*, *Z*1, and *Z*2 for rise and fall transitions. For both rise or fall transitions at node *Y*, when CLKB = 0, the intermediate node *X* is charged to voltage greater than half of the supply voltage. For a rise transition, the intermediate node *Z*1 is discharged to intermediate value, which charges the node *X* and the node *Z*2 is charged to supply voltage. Similarly, in the case of fall transition, node *Z*2 is discharged to an intermediate value which charges the node *X* and node *Z*1 is charged to supply voltage. In both cases, the warning signal is active low pulse. The delay of the warning generator is increased by 5× at a supply voltage of 0.7 V compared to the delay of warning generator at 1.2 V. The variability of the delay of warning generator is 8.7% and
32% at supply voltages of 1.2 and 0.7 V, respectively, with
3-sigma process variations (including both within-die variation and die-to-die variation).

B. Flip-Flop-Level Comparison

To quantify the area and power savings of the proposed warning FF compared to existing FFs, the canary FF [11], warning detection sequential [15], *in situ* monitor [20], and pulse-based timing error predictor [18] are implemented in industrial 130-nm CMOS technology. The comparison of different warning FFs is given in Table III.

1) Area: The proposed warning FF requires 1.95× area compared to the conventional FF. However, the proposed FF

TABLE III
COMPARISON OF WARNING FFS

Parameter	Proposed	Ref [11]	Ref [15]	Ref [20]	Ref [18]	DFF
Area (layout)	1.95×	3.45×	3.1×	3×	2.81×	1×
Number of buffers	1	2	3	1	2	NA *
Monitoring node	Master latch output	FF input	FF input	Master latch output	Master latch output	NA
Architecture	Charge sharing	Double sampling	Transition detection	Double sampling	Transition detection	NA
Clock to Q delay	1.05×	1.05×	1.01×	1.03×	0.97×	1×
Setup time(ps)	113.6	71	77	117.8	187.76	76
Hold time(ps)	-10.3	1.9	1.9	-13.2	-17	2
Clock power (μW)	4.62	8.51	4.05	8.49	4.04	3.97
Average Power (No warning)(μW)	11.5	20.45	15.6	19.2	15.8	7.32
Average Power (Warning)(μW)	14.5	19.21	17.1	19.2	17.3	NA
Peak Power (No warning)	1.14×	2.45×	1.12×	2.42×	1.82×	1×
Peak Power (Warning)	1×	1.66×	1.52×	1.81×	1.33×	NA

* Not Applicable

occupies 30% less area compared to [18]. This area savings are due to the requirement of less transistors to implement the warning generator. Moreover, the approach in [18] requires two delay buffers, one for warning margin and the other for transition detection unlike the proposed warning FF which requires only one delay buffer to provide warning margin. The area of canary FF [11], warning detection sequential [15], *in situ* monitors [20] and pulse-based predictor [18] is 3.45×,
3.1×, 3×, and 2.81×, respectively, compared to that of conventional FF as given in the second row of Table III. The large area overhead in [11] and [20] is due to the usage of double-sampling architecture. Moreover, the canary FF [11] requires extra delay buffers to account for the setup time of the main FF as it monitors at the input of FF.

2) Average Power: The proposed warning FF consumes 1.57× more power compared to the conventional FF. However, the proposed FF consumes 27% less power compared to the pulse-based timing error predictor [18] with 50% data activity at an operating frequency of 250 MHz. The reduction in power consumption is due to the less number of the transistors in the warning generator. Moreover, clock power of the proposed warning FF is just 1.16× to that of the conventional FF with 0% data activity. However, Canary FF [11] and *in situ* monitor [20] consume 2.14× and 2.13× more power, respectively, compared to the conventional FF with 0% data activity. This increased power consumption in [11] and [20] is due to the usage of more number of clocked transistors in the double-sampling architectures. In the proposed warning FF, the power consumption is more with the warning signal asserted compared to no warning condition. However, Canary FF consumes less power with warning signal asserted because of the reduced data activity at shadow FF compared to no warning condition. The power consumption due to the warning window generator of [15] is not taken into account for the comparison as it can be shared among warning FFs.

3) Peak Power: The peak power of the proposed FF is measured for different data to clock durations. When the data transition happens during the warning margin window, the peak power is higher compared to peak power due to a data transition before the warning margin window. The peak power of the different FFs is given in Table III. The double-sampling techniques [11], [20] consume higher peak power compared to transition detection approaches [15], [18]. The proposed FF consumes minimum peak power, out of the compared FFs, when the warning signal is flagged. When there is no warning signal, the proposed FF consumes 14% extra peak power compared to the conventional DFF. The warning detection sequential of [15] consumes 12% extra power compared to the conventional DFF.

TABLE IV
POWER SAVINGS OF ISCAS89 BENCHMARK CIRCUITS

Benchmark Circuit	Frequency (MHz)	%Replacement Rate	%Area Overhead	POFW (V)	%Power Savings
S13207	416	7	7.2	0.94	18
S15850	250	11.2	5.5	1	19
S38417	225	13.23	6.5	1	22

C. ISCAS89 Benchmark Circuits

The ISCAS89 benchmark circuits [27] are implemented to operate under worst case conditions with slow process corner, 125 °C temperature, and 10% drop in the supply voltage. The critical paths with slack less than 20% of time period are monitored with proposed FFs. The replacement rate is the ratio of number of proposed warning FFs to the total number of FFs in the design. Initially, the designs are simulated using a gate-level simulator (Synopsys VCS) under typical operating conditions with a nominal supply voltage of 1.2 V at a temperature of 25 °C. The timing libraries for different supply voltages are created using Cadence Liberate. The supply voltage of the design is reduced in steps of 20 mV until a warning signal is flagged. The POFW voltage is found by simulations for each design. From the simulation results, it is evident that by using the proposed FF, the designs can operate at reduced voltage under typical conditions. Power savings up to 22% can be obtained by employing the proposed FF in ISCAS89 benchmark circuits. The power savings obtained using the proposed FF in ISCAS89 benchmark circuits are given in Table IV.

VI. Conclusion

DVS and DFS with timing error monitors have become an effective way to reduce the worst case timing guard bands. This paper presented a low overhead warning FF, which monitors delayed master latch output to predict the timing vio- lations. The proposed warning FF consumes 27% less power at 50% activity factor and requires 30% less area compared to the timing error predictors available in the literature. A test chip is fabricated in industrial 130-nm CMOS technology to verify the effectiveness of the proposed warning FF in reducing the timing margin. Measurement of the test chip demonstrates that a design with the proposed warning FF can operate at
44% and 31% higher clock frequency at a supply voltage of 0.9 and 1.05 V, respectively, compared to the traditional worst case design in typical conditions. For a typical chip, the power consumption can be reduced by 36% compared to the conventional worst case design.

Acknowledgment

The authors would like to thank Europractice IC service for chip fabrication.

References

[1] T. Kuroda *et al.*, "Variable supply-voltage scheme for low-power high-speed CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 33, no. 3, pp. 454–462, Mar. 1998.

[2] J. W. Tschanz *et al.*, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1396–1402, Nov. 2002.

[3] J. T. Kao, M. Miyazaki, and A. P. Chandrakasan, "A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1545–1554, Nov. 2002.

[4] A. Drake *et al.*, "A distributed critical-path timing monitor for a 65 nm high-performance microprocessor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2007, pp. 398–399.

[5] D. Ernst *et al.*, "Razor: A low-power pipeline based on circuit-level timing speculation," in *Proc. Int. Symp. Microarchitecture*, Dec. 2003, pp. 7–18.

[6] S. Das *et al.*, "A self-tuning DVS processor using delay-error detection and correction," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 792–804, Apr. 2006.

[7] S. Das *et al.*, "RazorII: *In Situ* error detection and correction for PVT and SER tolerance," *IEEE J. Solid-State Circuits*, vol. 44, no. 1, pp. 32–48, Jan. 2009.

[8] T. Sato and Y. Kunitake, "A simple flip-flop circuit for typical-case designs for DFM," in *Proc. Int. Symp. Quality Electron. Des. (ISQED)*, Mar. 2007, pp. 539–544.

[9] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Adaptive performance compensation with in-situ timing error prediction for subthreshold circuits," in *Proc. Custom Integr. Circuits Conf. (CICC)*, 2009, pp. 215–218.

[10] M. Fojtik *et al.*, "Bubble Razor: Eliminating timing margins in an ARM Cortex-M3 processor in 45 nm CMOS using architecturally independent error detection and correction," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 66–81, Jan. 2013.

[11] H. Fuketa, M. Hashimoto, Y. Mitsuyama, and T. Onoye, "Adaptive performance compensation with in-situ timing error predictive sensors for subthreshold circuits," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 20, no. 2, pp. 333–343, Feb. 2012.

[12] B. P. Das and H. Onodera, "Warning prediction sequential for transient error prevention," in *Proc. IEEE Int. Symp. DFT VLSI Syst.*, Oct. 2010, pp. 382–390.

[13] M. Agarwal, B. C. Paul, M. Zhang, and S. Mitra, "Circuit failure prediction and its application to transistor aging," in *Proc. VLSI Test Symp. (VTS)*, May 2007, pp. 277–286.

[14] B. Rebaud *et al.*, "Timing slack monitoring under process and envi- ronmental variations: Application to a DSP performance optimization," *Microelectron. J.*, vol. 42, no. 5, pp. 718–732, May 2011.

[15] B. P. Das and H. Onodera, "Frequency-independent warning detection sequential for dynamic voltage and frequency scaling in ASICs," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 22, no. 12, pp. 2535–2548, Dec. 2014.

[16] T. Azam and D. R. S. Cumming, "Efficient sensor for robust low-power design in nano-CMOS technologies," *IET Electron. Lett.*, vol. 46, no. 11, pp. 773–775, May 2010.

[17] M. Wirnshofer, L. Heiß, G. Georgakos, and D. Schmitt-Landsiedel, "A variation-aware adaptive voltage scaling technique based on in-situ delay monitoring," in *Proc. IEEE Int. Symp. Design Diagnostics Electron. Circuits Syst. (SDDECS)*, Apr. 2011, pp. 261–266.

[18] L.-Y. Chiou, C.-R. Huang, and M.-H. Wu, "A power-efficient pulse-based in-situ timing error predictor for PVT-variation sensitive circuits," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2014, pp. 1215–1218.

[19] V. Huard *et al.*, "Adaptive wearout Management with *in-situ* aging monitors," in *Proc. IEEE Int. Rel. Phys. Symp.*, Jun. 2014, pp. 1–11.

[20] M. Saliva *et al.*, "Digital circuits reliability with *in-situ* monitors in 28 nm fully depleted SOI," in *Proc. Design, Autom. Test Eur. Conf. Exhib. (DATE)*, 2015, pp. 441–446.

[21] A. Benhassain *et al.*, "Timing in-situ monitors: Implementation strategy and applications results," in *Proc. Custom Integr. Circuits Conf. (CICC)*, 2015, pp. 1–4.

[22] W. Shan, X. Shang, L. Shi, W. Dai, and J. Yang, "Timing error prediction AVFS with detection window tuning for wide-operating-range ICs," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, to be published.

[23] Y. Zhang *et al.*, "iRazor: 3-transistor current-based error detection and correction in an ARM Cortex-R4 processor," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2016, pp. 160–162.

[24] X. Shang, W. Shan, L. Shi, X. Wan, and J. Yang, "A 0.44 V-1.1 V 9-transistor transition-detector and half-path error detection technique for low power applications," in *Proc. A-SSCC*, Nov. 2017, pp. 205–208.

[25] I. Kwon, S. Kim, D. Fick, M. Kim, Y.-P. Chen, and D. Sylvester, "Razor-lite: A light-weight register for error detection by observing virtual supply rails," *IEEE J. Solid-State Circuits*, vol. 49, no. 9, pp. 2054–2066, Sep. 2014.

[26] M. Omaña, D. Rossi, N. Bosio, and C. Metra, "Low cost NBTI degradation detection and masking approaches," *IEEE Trans. Comput.*, vol. 62, no. 3, pp. 496–509, Mar. 2013.

[27] (Nov. 2017). *ISCAS'89 Benchmarks*. [Online]. Available: http://www.pld.ttu.ee/maksim/benchmarks/iscas89/verilog/

Clock-Gating of streaming applications for energy efficient implementations on FPGAs

1. K. Mounika, M.Tech student, Mallareddy College Of Engineering

2. D. Rajendra Prasad, Assistant Professor, Mallareddy College Of Engineering

Abstract— **This project deals with reducing the dynamic power in streaming applications by using clock-gating. Streaming applications consists of broad class of computing algorithms used in areas such as cryptography, digital media coding ,communications, signal processing, video analytics , network routing etc.,. The power saving can be gained by selectively turning off or switching off parts of the circuit when they are temporarily inactive. So, that the switching state power can be reduced. In this the clock gating methodology is introduced in dataflow designs that are automatically included in the synthesis stage of the high level dataflow design flow. This concept describes an approach for developing energy optimized run-time reconfigurable design which is a benefit from clock gating. For this, FPGAs are highly desirable due to its reconfigurable (or) re-programmable nature, flexibility. It uses clock gating strategy for reducing dynamic power. The main aim of this project is to reduce dynamic power dissipation without much affecting performance and the experimental results also show that the applications synthesized on FPGA platforms shows that power reductions can be achieved with no loss in data throughput. By using this approach we can achieve low power , delay, area and low noise levels**.

Index Terms— **Clock-gating, data flow , dynamic power, high level synthesis**

I. INTRODUCTION

In general, power consumption is one of the major challenges in VLSI design performance. For a silicon device there are two components of power dissipation. They are static power dissipation and dynamic power dissipation. Static power dissipation is due to the leakage current produced within the transistor and by the ambient temperature. Dynamic

Dynamic power consumption can contribute upto 50% of the total power consumption. Inorder to reduce unwanted power dissipation additional circuit is added into the design that effectively clocks the gate. Power dissipation increases linearly with frequency due to highly influence of parasitic capacitances. To counteract this effect, ASIC designers have employed clock gating from past few years[1],[2],[3]. Clock gating is nothing but a power saving feature used in semiconductor devices which enables switching off or turning off circuits. Many silicon devices use clock gating to switch off controllers, parts of processor, bridges and buses inorder to reduces the power dissipation.

power dissipation is due to switching of transistors and by losses of charges being moved along wires.

Reducing power has also another benefits. They are it needs less stringent to cool the device, improved life of device or battery and low power costs. Due to these reasons, the power also frequently affects the choice of computing platform right at outset. For instance, Field-Programmable Gate Array (FPGA) imply higher power dissipation per logic unit by comparing with ASIC. This paper shows the impact of chosen technology on the architecture, but do not describe how to reduce the power at abstraction level of the design.

The major difference between FPGA and ASIC is ASIC is fixed implementation, that means these are pre-defined for a specific task, where as FPGA can be reprogrammed ON- site. FPGA eliminates non- recurring Engineering costs and thereby reduces time to market. Therefore FPGAs are highly desirable to implement digital systems due to their programmability , low end product life cycle, flexibility and all these makes ideal for small and medium applications. FPGAs are slower and less efficient due to added circuitry that is used to make flexible when compared to fixed implementation(ASIC).

Dynamic dataflow designs, for instance the RVC- CAL language possess interesting properties that are used for reducing the power without affecting the behavioral characteristics and the construction of the application. In RVC- CAL language, every actor can execute the processing tasks, executions may be disabled by the input blocking reads and communication between the actors can occur only by the order of storing lossless queue. As a result, an actor may be stopped for a period of time if its processing tasks are idle(or) output buffers (queues) are full without affecting overall throughput and semantical behavior of the design.

II. RELATED WORK & METHODOLOGY

In this paper, the work and methodology described below are based on Xilinx FPGA, which can be used to support the other architectures. A system which works or process on continues stream of bits is said to be streaming application. Cryptography, digital media coding, network routing, video analytics etc., are the examples. In clock gating scheme(CG),

the clock is given to those modules that are working at that instant. This clock-gating support the adding of additional logic to the existing synchronous circuit inorder to prune the clock tree, thereby disabling the portions of the circuit that are not in use. In an architecture, the additional device is called clocking circuit which is inserted before the data path unit which provides the clock inputs for working or active modules only. Hence it reduces dynamic power.

Related to our work, S.C Brunet, E. Bezati, C. Alberti M. Mattavelli, E. Amaldi, and J.W.Janneck proposed a multiple clock, domain-design methodology inorder to reduce the power consumption of dataflow programs. Their design motive was to optimize the mapping of the application along with meeting the desired design requirements. The optimization is gained by assigning optimized clock frequency to reduce power consumption.

Present FPGA support various clock-gating strategies and every manufacturer creates its own IP [5] for managing these approaches. The methodology explained here is based on primitives used specific to Xilinx FPGA architectures. Anyway these are modified inorder to support other FPGA vendor primitives.

The execution of a dataflow program contains a series of action firings [2],[6], which can be correlated to one another in a graph-based representation using an approach called Execution Trace Graphing (ETG). In this graph each node represents an action firing and directed arcs represent data (or) control dependency between two various action firings. ETG is a directed and acyclic graph.

By using weighted ETG more optimized buffer size can be gained [8],[9],[10]. In this graph, each firing action is represented with its timing information, hence transform it into weighted graph. It consists of two parts

A. Clock-gating strategy
B. Clock enabling circuit

A. Clock – gating strategy:

The following fig.1 shows the clock gating scheme or strategy. The blocks it contains mainly queue, clock enabling circuit and actor. Actor will be chosen based on the application taken. In this the clock enabling circuit controls the clock of the actor. Whenever the output buffer of the block becomes full then the clock should be turned off inorder to show that the block is in idle state. Switching off the block does not show any effect on the throughput because it is in idle state, there by reduces the power dissipation.

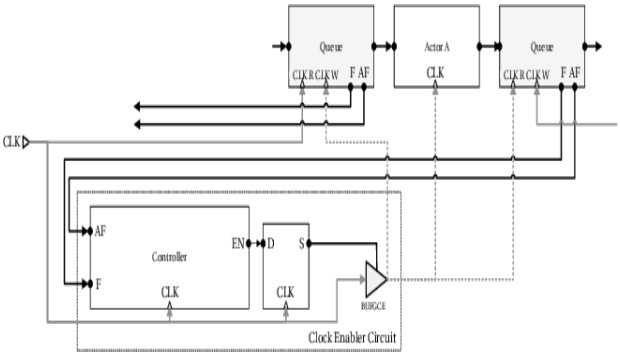


Figure.1 Clock Gating Strategy

Generally RVC-CAL dataflow designs are used for the behavior description, that can be applied to systems that represent execution of process that communicate with asynchronous FIFO buffers. Here the queue blocks should have lossless communication when the actor is clock gated with asynchronous buffers and the design will have different input clock domains.

In the figure shown each queue will have two clocks as inputs, CLKW and CLKR. CLKW is for taking data and CLKR is for producing data. And there are two output ports for the queue. They are Almost Full(AF) and Full (F). From the queue1, the data is given as input to the actor and from the actor the output data is collected by queue2, from this queue2 is readout. When the output of the actor is full, then queue1 will be stopped. To the queue1 the enabling circuit will provide the clock, which will turn off the actor when it is full, hence saving the power. From the fig.1 shown above, the input to the actor is connected to the clock enabling circuit. The clock buffer BUFGCE input clock should be connected to a flip flop to achieve glitch-free clock gating [7].

B. Clock enabling circuit:

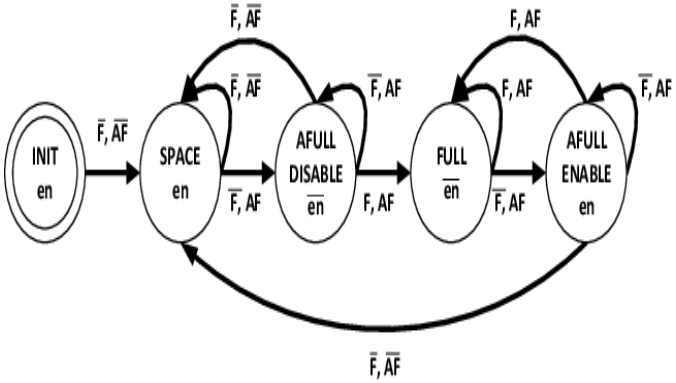


Figure.2 Clock enabling circuit implemented as Finite State Machine

The clock enabling circuit is shown in above fig.2. The shown fig. is implemented as a finite state machine (FSM).

FSM is having a clock, a reset, input Full(F), input Almost Full (AF) and an output enable (EN). The Almost Full AF input becomes active high only when there is one space left to full in its FIFO queue [5]. The above shown FSM has 5 states.

S= {INIT , SPACE , AFULL_DISABLE , FULL , AFULL ENABLE }

The clock-enabling controller circuit starts with INIT state and maintains the ENABLE EN output at active HIGH until F and AF become active LOW. The active high ENABLE EN is maintained during SPACE state. As soon as a queue becomes full, the state will change to AFULL_DISABLE. Here the queue gets disable. In this case, EN output becomes active LOW.

An approach is used in this state as BUFGCE a clock buffer disables the output clock from high-to-low level edge. The enable clock entering BUFGCE must be synchronised to input clock. When a token is taken from the queue, it goes to the AFULL_ENABLE and activates the clock. Then based on the output of buffer Full F (or) Almost Full AF , the state changes to Full (or) to SPACE state.

The user can choose the actor to be clock gated by using mapping configuration. For this, an attribute is given to each actor. The outputs of FIFO queues F and AF are connected to the clock enable controller circuit , if an actor is selected for clock- gating. Output of the queues can be connected directly to a queue (or) through fanout. In the first case, actor output is directly connected to queue without fanout. In the second case, the controller results are connected to other port. In this case one of the fanouts in the queue is full and the fanout should command the actor , but not to produce a token.

III. EXPERIMENTAL RESULTS

This section gives the power reduction gain of the previously mentioned methodology, which is evaluated by applying it to a video decoder design. There are many RVC-CAL applications for dataflow programs [11]. INTRA MPEG4 simple profile decoder is one of the applications. Due to the limitation of number of clock buffers in Xilinx FPGA the design selected was redesigned to result in 32 actors.

MPEG4 video coding is characterized by its scalability and high flexibility. It is a method of compression of audio and the visual digital data. Here the Intra MPEG 4 Simple Profile (SP) description consists of 32 actors. It is a 4:2:0 decoder consists of 8 blocks. Out of these blocks, 4 blocks are luminance (Y) and the other are chrominance U and V, for each two blocks. The parser block contains the bitstream and variable length decoding process, in which U, V and Y are used for texture implements (Tex Y, U,V). The texture decoding consists of variable length coding (VLC), inverse scan test of DCT, inverse quantization, inverse DCT. Whereas, MOT Y, U, V are used to realize the motion compensation stage. Due to the nature of experiments, the MOT stage consists of only one residual error actor. For each queue in the decoder, the minimum queue size is determined [4] by using TURNUS profiler.

For hardware experiment and evaluation, VC707 FPGA kit was used. The code was generated by xronos and was synthesized by Xilinx XST synthesizer. Inorder to produce a netlist, synthesis, routing and placing were applied. The final netlist generated was simulated inorder to extract the Switching Activity Information File(SAIF) of the design. After that, the Xilinx power analyzer was used to get power consumed by using design constraints, design netlist and SAIF inputs also.

The following Table I shows the synthesis results of Intra MPEG4 SP decoder with and without clock-gating.

Logic utilization	Non clock gated	Clock gated	Available
Slices	9214	12776	607200
DSPs	18	18	2850
BRAMs	7	7	1030
LUTs	21499	25126	303600
Max freq.	109	109	-

Table I : Synthesis results with and without clock-gating

A safe option when finalizing your figures is to strip out the fonts before you save the files, creating “outline” type. This converts fonts to artwork what will appear uniformly on any screen.

The above table I shows that clock gated decoder use more slices than non-clock gated. Here the clock gating needs only 15% more than LUTs. A 50 MHz clock is given as synthesis constraint. The Table II shown below gives the power consumption of Clock Gating strategy. In this, two tests are considered. Clock gating enabled and disabled [5], when decoding at maximum output.

Clock gating	Disabled(mW)	Enabled(mW)
Actors clock	58	43
Logic	25	24
Signals	42	41
Clocks	94	80
leakage	242	242
total	403	387

Table II : power consumption when clock gating enabled and disabled

From Table II, the actor clocks show only the power consumption of the actor. Where as, the clock will consist of 50 MHz clock net enabling of clock nets. As a result, the actor clocks consume 26% less power due to clock gating and due to decoder running at high speed, the activation rate of signals and logic result in decrease of 4% in total power.

As from Table I, the maximum decoder output rate is 350 frames/second, for QCIF image of 176x144 pixels. Here the decoder is throttled, such that to decode only 30 images/second for two resolutions. They are CIF(384 x 288 pixels) and QCIF.

The following fig.4 shows (a) the actor power consumption and (b) shows activation rate of actor for each actor clock. The activation rate shows that some of them has an activation rate less than 10%. From this, the power consumption on clocks has reduced by 53.7% for QCIF and 47.6% for CIF resolution. The decoder consumes 54mW less for CIF resolution and 59mW less for QCIF resolution when compared to overall power consumption. From 31 actors, 15 will be always ON, this implies that the 15 actors will never fill up their output. Further, the actors which are not needed for this methodology should find out and eliminate the installation of unwanted additional logic.

In the case of **power saving efficiency and bandwidth,** the decoder will be throttled from 0 to 90% [5] by simulating channel with different consumption rates. It is a case where clock gating is applied for a general application. It is shown in fig.3 below.

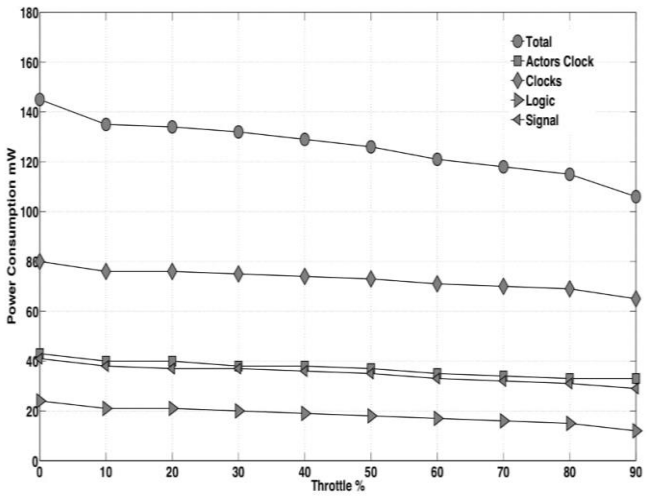
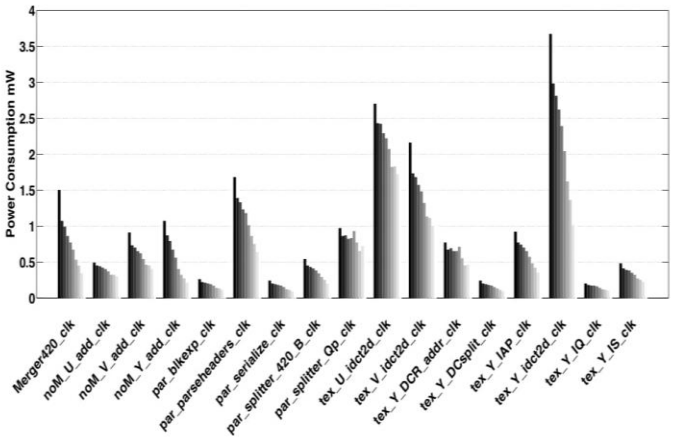


Figure.3 power consumption of clocks, signals, logic and total dynamic power consumption of INTRA MPEG4 SP decoder

From the figure shown above, it is depicted that the total dynamic power has reduced from 145mW to 106mW, a total reduction of 27% power.

The dynamic power of clock gating has been reduced by 34% when compared to non-clock gating. From fig.4 it is shown that, the 15 actors data has been removed due to their activation rate being higher than 99%.



.

Figure. 4(a) Actor clock power consumption

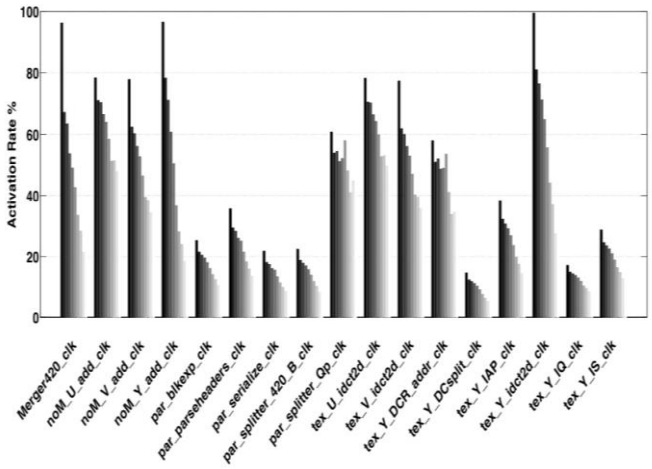


Figure.4(b) Actor clock activation rate

The clock activation rates of actors will be decreased with increase of throttle(except for two cases par_splitter_QP_clk & tex_Y_DCR_addr_clk where power increased slightly). Here, the decoder used is [5] YUV 420. When the power reaches 60%, the chrominance decoding remain active where as, the decoder throttles luminance. It is also occurred during a behavioral simulation in ModelSim.

IV. CONCLUSION

This clock-gating (CG) methodology can be applied to any application inorder to reduce power consumption and reduces the effort during design process at dataflow level. The clock gating(CG) logic is introduced at synthesis stage of HLS design flow. The important component in clock gating is clock enable controller, which is used to give the additional clock. The results show that, when the main block is inactive due to the disabled clock it reduces the switching power. This methodology is very useful where the power dissipation is major challenge.

This method is a simple and effective method inorder to recover power from an idle state. Further developments are necessary inorder to develop the tools used for complex applications onto the limited number of clock domains for more implementations.

V. REFERENCES

[1]. Massoud Pedram, "Power minimization in ic design: Principles and applications," ACM Trans. Des. Autom. Electron. Syst., vol. 1, no. 1, pp. 3–56, Jan. 1996.

[2]. QingWu, M. Pedram, and XunweiWu, "Clock-gating and its application to low power design of sequential circuits," Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on, vol. 47, no. 3, p. 415–420, Mar 2000.

[3]. G.E. Tellez, A. Farrahi, and M. Sarrafzadeh, "Activity-driven clock design for low power circuits," in Computer-Aided Design, 1995. ICCAD-95. Digest of Technical Papers., 1995 IEEE/ACM International Conference on, Nov 1995, pp. 62–65.

[4]. M. Canale, S. Casale-Brunet, E. Bezati, M. Mattavelli, and J. Janneck, "Dataflow programs analysis and optimization using model predictive control techniques," Journal of Signal Processing Systems, pp. 1–11, 2015.

[5]. Endri Bezati, Simone Casale-Brunet, Member, IEEE Marco Mattavelli, and Jorn W. Janneck, Member, IEEE "clock-gating of streaming aplications for energy efficient implementations on FPGAs" DEC-2015.

[6]. H Mahmoodi, V Tirumala shetty, M Cooke and K Roy," Ultra Low-Power Clocking Scheme Using Energy Recovery and Clock Gating," *in IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 17, no. 1, pp. 33-44, Jan. 2009.

[7]. Xilinx, Analysis of Power Savings from Intelligent Clock Gating, August 2012, XAPP790.

[8]. S. C. Brunet, M. Mattavelli and J. W. Janneck, "Buffer optimization based on critical path analysis of a dataflow program design," *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, Beijing, pp. 1384-1387,2013.

[9]. B. Ghavami and H. Pedram, "High performance asynchronous design flow using a novel state performance analysis method, *Computational Electrical Engineering*, vol.35, no.6, pp. 929-941, November 2009.

[10]. S Suhaib, D Mathaikutty, and S Shukla, "Dataow architectures for GALS," *Electron. Notes Theor. Comput. Sci.*, vol. 200, no. 1, pp. 33–50, 2008.

[11]. "Open RVC-CAL Applications," 2014, http://github.com/orcc/orc-apps, accessed 25-February-2014.

TRANSITION INVERSION BASED LOW POWER DATA CODING SCHEME FOR BUFFERED DATA TRANSFER

Suhasini Parunandi
PG Student
psuhasini11@gmail.com

Malla Reddy College of Engineering (MRCE) Dulapally ,Telangana ,India

Abstract

In this work the authors propose a data coding protocol that leads to power reduction for block data transfer in off-chip buses. I/O pads driving off-chip buses contribute to a major portion of power dissipation in chips. Also, block data transfer is preferred in most systems like caches, DMA etc. In this proposed work, the prior knowledge of the block of data to be transmitted, when it is stored in the buffer, is exploited in a serial fashion to reduce transitions on every bus line. Statistical analysis shows up to 31.9% reduction in transitions. Benchmark results show that it leads to 29% reduction in power consumption. The technique provides added error detection on the lines of parity bit technique, with similar average error detection capability.

1. Introduction

Increased integration and higher operating frequencies compound the problem of power dissipation in VLSI chips. One of the main contributors to power consumption is switching activity on the high-capacitance lines of an interconnection system, especially off-chip buses. Buses constitute an important resource for addressing and data transfer in the implementation of most electronic systems. The fact that the power consumed at the I/O pads accounts for a significant fraction of the total power consumed in VLSI systems has been independently established by many researchers [1-3].

Error detection is also of equal importance, as buses are more prone to error due to temperature variations, interference from neighboring sources, and ageing etc. Thus a low power data coding scheme supporting error detection is desired. Existing techniques make use of the data as it is placed on to the bus. This leads to a delay which can severely affect the operating frequency of the circuit. The proposed encoding technique, which is based on serial transition inversion, works on blocks of data for power reduction in off-chip buses. Numerous bus transmission protocols deal with blocks of data rather than individual data words. This is evident in DMA transfers and cache lines which are widely used in computer systems. In block data transfer the latency in data coding can be hidden since the block will be transmitted only after it is filled up. This brings the pipeline approach to the coding technique. Since Off-chip buses consume more power, keeping the encoding circuitry before the I/O pad reduces power consumed by it. The proposed technique is also called transition inversion in the following sections

2. Related work

One of the most often cited encoding methods is the bus-invert method [1]. Bus-invert selects between the original and the inverted pattern in a way that minimizes the switching activity on the bus. The resulting patterns together with an extra bus line (to notify whether the address or its complement has been sent) are signaled over the bus. Musoll et al. proposed the working zone method [3]. Their method takes advantage of the fact that data accesses tend to remain in a small set of working zones. Other encoding techniques include Asymptotic Zero-Transition Encoding [2], Gray coding (mostly for addresses), and other application specific encoding techniques [9,10]. Most of the existent techniques make use of the repeating patterns in address buses to reduce address bus transitions [6,8,9]. Frequent Value encoding (FVE) is another technique proposed in [13] also results in a significant reduction in transitions, but has not been considered here, due to the overhead involved.

There is no existent literature on bus coding methodologies for block data transfer, other than Serial Bus Invert [14], which encodes blocks of data rather than individual data words. Also a transition inversion

based encoding technique [12] was developed for serial buses that serve as a basis for the proposed technique. The technique proposed in this work is compared with Serial Bus Invert, serial gray coding [15], and transition signaling [11], which is the reverse of gray coding. The last technique mentioned is similar to the algorithm proposed by us, but differs in the decision making process.

3. Proposed technique

In block data transfer, data is generally loaded onto a buffer and then is put on the bus. Each line in the bus is a serial line that will transmit one particular bit position of all data words that are put on the bus. Before transmission, the number of transitions on a line is counted. This can be done by a simple XOR gate between consecutive bits and counting the number of '1's. If the number of transitions is more than half the number of data words, the transitions states between the bits can be inverted. Each transition is made as a non-transition and vice versa. If not, the bit stream is transmitted as such. In case transition inversion is needed, the scheme operates by observing the transition states between any 2 bits and setting the encoded second bit to be the same as the previous encoded bit if there is a transition. If there is no transition, the previous encoded bit is inverted. The decision bit signifying transition inversion is transmitted before transmitting the encoded data. This has to be done on all lines. Since the data is sent as a block, the extra bit on each line will signal for all the data words.

The transitions in the bit stream, transmitted on a line, can be reduced by the aforementioned scheme. Each line is processed independent of each other. If there is a need for transition inversion, then the following steps are followed to obtain encoded data. Let the data bit that is to be transmitted next be b_d and previous data bit be b_{dp}. The previous transmitted bit is b_{tp}. The next transmitted bit will be

bt = btp if bd ≠ bdp
= !btp if bd = bdp

In receiver the reverse logic needs to be applied. When the bit stream has been signaled as modified, then the following steps are followed to decode data. The previous and current received bits are assumed to be 'brp' and 'br' respectively. The previous decoded bit is assumed to be bdp. The current received bit will be

bd = bdp if brp ≠ br
= !bdp if brp = br

The encoding and decoding is done on the fly, to reduce performance losses. For example in the bit stream 10101011, the number of transitions is 6. Thus this stream is to be modified according to the algorithm described above. The first bit is transmitted as such, without any change. This is described in Table 1. The encoded data has only one transition.

Table 1: Sample coding process

Key: NT – No Transition, T – Transition, NC-No Change

Bit No.	1	2	3	4	5	6	7	8
Bit stream	1	0	1	0	1	0	1	1
Transition State	NC	T	T	T	T	T	T	NT
Encode state	NC	NT	NT	NT	NT	NT	NT	T
Encoded Bit stream	1	1	1	1	1	1	1	0
Decode state	NC	T	T	T	T	T	T	NT
Decoded bit stream	1	0	1	0	1	0	1	1

A. Statistical Analysis of the Algorithm

A statistical analysis of the serial transition inversion (STI) algorithm has been carried out in two independent methods taking buffer depths of 8, 16, 32, and 64. The first analysis considers all combinations of the word and determines the original and modified transitions in the datasets. It is essentially a brute force approach. The data considered was a uniform distribution of all possible data patterns that is likely to be transmitted over the bus. For example, considering a buffer depth of 8, the number of possible data patterns of one bit stream is 256. The number of transitions in these data patterns was calculated along with the number of transitions in the data pattern after being modified, using the proposed algorithm. The second analysis was an analytical one. For an N-bit system, where the transitions are taken between the consecutive bits, maximum of (N-1) transitions are possible in the bit stream. Number of possibilities of 'i' transitions= $^{(N-1)}C_i$. Let T_{org} and T_{mod} be the number of transitions in the original data patterns and the number of transitions in the modified data patterns respectively. These entities can be calculated as follows: Total number of transitions T_{org}

$$\sum_{i=0}^{N-1}\left((N-1)_{C_i} * i\right) \quad (1)$$

Transition inversion is done when the number of transition is more than or equal to N/2. The number of transitions in the modified data will be (N-1-i) for 'i' transitions in the original data.

$$T_{mod} = \sum_{i=0}^{\frac{N}{2}-1}\left(N-1\right)_{C_i} * 2 * i + \sum_{i=\frac{N}{2}}^{N-1}\left(N-1\right)_{C_i} * 2 * N(-1-i) \quad (2)$$

The average reduction can be calculated by taking the difference between the number of transitions in the modified data patterns, given by equation (2) and unmodified data patterns, given by equation (1). The reduction figures would be smaller if the decision bit is also considered. The results obtained by both the methods agree with each other and are shown in Table 2.

Table 2: Statistical Percentage Reduction

Word Length	8 Bits	16 Bits	32 bits	64 bits
% Reduction in transitions	31.25	20.95	13.54	9.78

B. Power Analysis of the Algorithm

The overall power reduction consists of the power reduction achieved by the transition reduction minus the power consumed by the extra circuitry. Unmodified dynamic power consumed by I/O pads is given by

$$p_{org} = \frac{1}{2} V_{dd}{}^2 C_T f \alpha \qquad (3)$$

Where, Vdd, f, Ct, α represent drain voltage, frequency of operation, line capacitance, switching activity respectively. If the power dissipation of the extra circuitry required for the coding process is taken into account then the equation given above has two extra terms on the right hand side, the encoder and decoder power dissipation respectively.

C. Performance Analysis of the Algorithm

The transition inversion algorithm needs an extra bit to be transmitted before the start of the block of data on all lines. This leads to a decrease in bandwidth utilization. For a system with buffer depth of 8, 9 bits are transmitted on one line. Therefore a frequency increase of 9/8 will be needed to maintain the same bandwidth utilization. The corresponding power consumed by I/O pads will increase linearly.

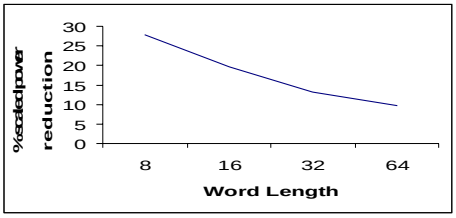


Figure 1: Effects of frequency scaling with word length

A performance metric is defined to take into account the scaling of the frequency and the reduction in transitions, and is calculated as their product. The variation of this parameter with word length is shown in Figure 1. Bus invert leads to a reduction in bandwidth since it poses a delay in putting the encoded data. By following a depth based approach where most delays are hidden, bandwidth need not be reduced.

D. Error Detection Analysis of the Algorithm

The proposed algorithm's propensity towards reducing the number of transitions can be used for detecting errors. This can be done by determining if the number of transitions in the received bitstream is more than half the bitstream length. If this count is more than half the bitstream length, the incoming data is incorrect. The proposed technique is compared with parity bit technique, as both have similar overhead i.e. addition of one bit to the bitstream. The parity bit detects all odd bit errors, but misses even bit flips, whereas, transition inversion can detect a certain percentage of any number of bit errors. Error analysis has been done by considering all combinations of the given word length that are transmitted over the bus. For transition inversion coding, the possible combinations of the word are those in which the number of transitions is less than 4. All the combinations of bit errors right from one bit error to 8 bit errors have been checked for both the proposed technique and parity bit technique. The result of this analysis is shown in Table 3.

Table 3: Error Detection Analysis

No. of Bit errors	% of errors detected	
	Parity Coding	Proposed Technique
1	100	31.25
2	0	44.64
3	100	52.68
4	0	55.71
5	100	52.68
6	0	44.64
7	100	31.25
8	0	0

If all the bits are in error, then neither technique can detect the error, as in the proposed technique if all bits are flipped, the number of transitions remains the same. Calculation of statistical averages over the entire range of bit errors shows that the proposed technique and parity bit technique both have the same value of 50.2%. The average is calculated as the ratio of total number of errors detected to the total number of errors possible on the line. Thus the proposed technique can be used as a hint to upper layers of communication that an error has occurred since it cannot reliably detect all errors.

4. Implementation strategies

In most block systems, the data buffer is present just before the transmission part. The core logic puts the data inside the buffer from one side and the transmission happens from the other side.

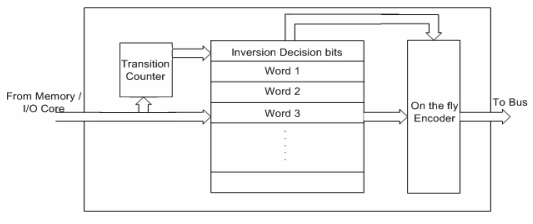


Figure 2: High Level Architecture

The transition counting happens when the data is being filled up in the buffer. The transition inversion decision is made depending on the count of the transitions and encoding done based on it. The bit stream is encoded on the fly as the data is put on the bus, as shown in Figure 2. In the receiver the decoder has to decode the incoming bit stream and recover the original data.

A. Decision Circuit and Encoder

The decision circuit is a simple XOR gate between consecutive bits of the input bit stream as shown in Figure 3. The counter needs to count only up to half the number of maximum transitions.

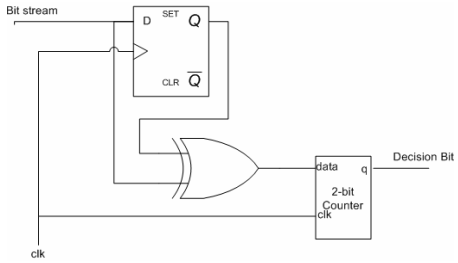


Figure 3: Decision Circuit (Transition Counter)

This circuit can also be implemented with double edge triggered circuits to further optimize at the encoder stage.

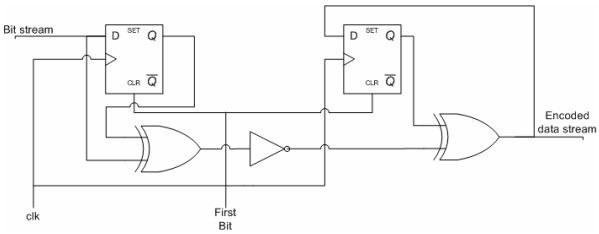


Figure 4: Encoder Circuit

The transition counter works in parallel to buffer loading, and is thus masked. The on the fly encoder is shown in Fig 4.

B. Decoder

The decoder shown in Fig 5 performs XOR between consecutive bits to determine transition state, inverts the received bit if required to recover the data.

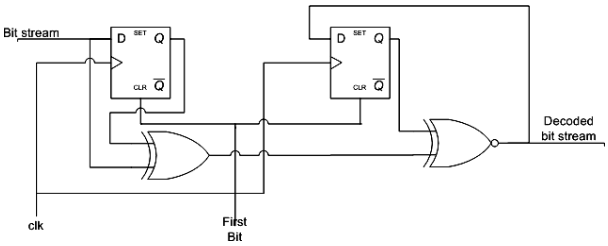


Figure 5: Decoder Circuit.

C. Complexity Analysis

The main components of both systems are the decision circuit, encoder and decoder.

Decision circuit: With increase in bus width, the space complexity increases exponentially ($O(N^2)$) for bus invert decision circuit. A comparable parameter in the proposed technique is buffer depth which leads to a linear increase (O(N)) in circuit complexity. Time complexity increases linearly (O(N)) in bus invert, while in the proposed technique it is constant(O(1)).

Encoder and Decoder: With increase in bus width, the bus invert encoder/decoder circuit complexity increases linearly (O(N)). For the proposed technique, the circuit complexity is constant (O(1)). Time complexity is constant (O(1)) in both bus invert and the proposed technique.

5. Experimental results

For experimental analysis, the algorithm was applied on random image data and SPEC2000 benchmark binaries.

Random Image Data:

For this analysis of the algorithm seven images were taken and their RGB values were ran through the algorithm. The images were a mix of both smooth and detailed features. The results are tabulated in Table 4. These do not include the power dissipated by the encoder and decoder circuitry.

Table 4: A comparison of transition reduction for Bus invert and the Proposed Technique

#	Original no. of transitions	Bus Invert Coding		Proposed technique	
		transitions	% reduction	transitions	% reduction
1	120160	86296	28.18	72212	39.9
2	127770	94776	25.82	85454	33.11
3	74666	61746	17.3	53502	28.34
4	165678	119578	27.83	119908	27.62
5	111909	81645	27.04	70978	36.58
6	66189	49251	25.59	46769	29.34
7	159620	121466	23.9	114163	28.48

It is clear from the above table that STI performs much better than bus invert.

With SPEC2000 benchmarks:

SPEC2000 benchmark binaries traces were run with the proposed technique and compared with bus invert and gray coding. The 26 binaries were run with varying the buffer depth and bus widths with the values 8,16,32,64. The averages for a given combination of bus width and buffer depth were taken and have been plotted. The results for the proposed technique and bit invert are showed in Figure 6 and 7 respectively.

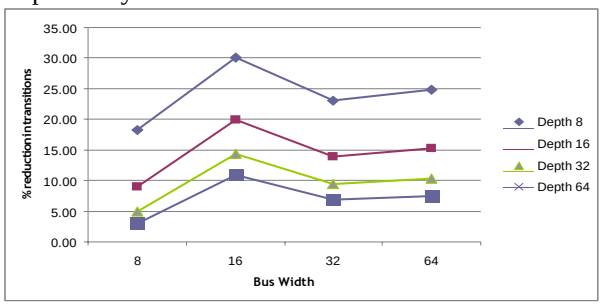


Figure 6: Transition reduction in Transition Inversion

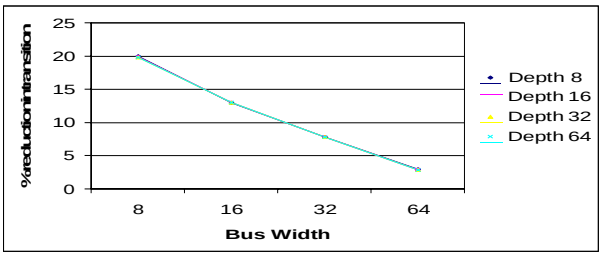


Figure 7: Transition reduction in Bus invert

It can be observed that the buffer depth does not make any changes to bus invert. Also with increase in buffer depth, the transition reduction reduces for the proposed technique. A similar observation can be made for bus invert when bus width is increased.

With increasing bus widths in present VLSI systems, the proposed technique will perform better. The increase in buffer depth leads to a lesser reduction. It can be offset by splitting the block into sub-blocks of smaller depths. Depending on the system, a compromise between power reduction and bandwidth utilization can be found. The benchmark files were also run using Gray Coding technique which is the reverse of transition signaling. The results are shown in Figure 8.

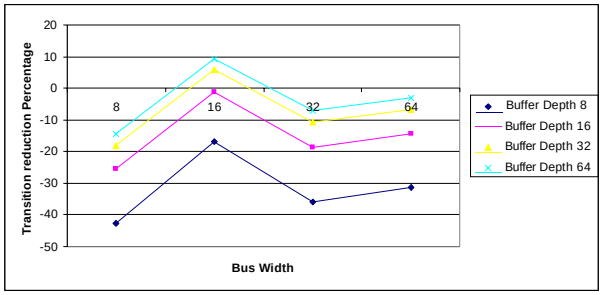


Figure 8: Transition reduction in Gray Coding

It can be seen that there is not much reduction in transitions, with some of the data points showing an increase in the number of transitions. Gray code does not show any reduction in transition since it is an N-bit to N-bit mapping. Whatever data tuples are in input set are exactly the same in the output set. So it is not possible to get any benefits out of Gray code. The same is the case with transition signaling.

B. Overall Power Analysis

The proposed system was designed in RTL and analyzed with Synopsys synthesis tools. A bus operating at 100MHz was assumed with its I/O voltage levels at 3.3v. The internal circuitry was modeled on 180nm process technology. The circuitry was simulated by feeding the SPEC2000 benchmark trace files as input for a buffer depth of 8. The power consumed by the circuitry is shown in Table 5.

Table 5: Power dissipation of encoding/decoding circuitry for transition inversion

Circuitry	Decision circuit	Encoder	Decoder
Power consumed	28.7µW	28.9µW	28.6µW

Assuming the parameters stated above and the activity factor for the benchmarks to be 0.5, the power was found to be 27.23mW. The reduction in power consumption is linearly dependent on the activity factor reduction. A reduction of 30% activity leads to a reduction of power by 8.17mW. The total power consumed by the extra circuitry is 86.2 µW leading to a

net power reduction of 8.08mW which corresponds to 29.7 % reduction in power.

C. Overall Delay Analysis

The proposed technique does not involve circuitry of multiple stages thus leading to less delay. The delay performance of the proposed technique and bus invert in terms of propagation speed is compared in Table 6.

Table 6: Comparison of encoder delay in Bus Invert and Transition Inversion

Technique	Proposed Technique	Bus Invert
Delay	1.2ns	3.3ns

For calculating the delay due to the proposed technique only the encoder is considered. The decision circuit is not taken as it will be part of the buffer loading delay and will not contribute to encoding. The decision circuit delay was found to be 0.2ns since it involves only the XOR gate. The flip flops delays will be masked by the sequential loading thus giving a pipelined approach. For the bus invert technique, the decision circuit delay is also taken into account because encoding has to be complete before the next data word arrives. Thus before the next data arrives the counting of the hamming distance should have been done and the data encoded. Overall the encoding delay of the proposed technique is considerably lesser compared to that of the bus invert.

6. Conclusions

In this paper an encoding technique has been presented that reduces power dissipated on off-chip data buses for block data transfer. The technique involves inverting the transition states on every line of the bus if the transitions exceed the number of non-transitions. The modification status is signaled as an extra word, thus avoiding the use of an extra line. The average reduction obtained in terms of transitions is 31.9% while the net power reduction after the extra power circuitry is taken into account is 29.7%. This is achieved without using an extra bus line. The compromise is in bandwidth utilization which can be adjusted by choosing a proper block length. This technique can also be applied to synchronous serial buses. The presence of a parity bit like error detection mechanism in addition to low power gives it an additional advantage.

7 References

[1] M. R. Stan, W. P. Burleson. “Bus-Invert Coding for Low Power I/O”, IEEE Transactions on Very Large Integration Systems, Vol. 3, No. 1, pp. 49-58, March 1995.

[2] L. Benini, G. De Micheli, E. Macii, D. Sciuto, C. Silvano. “Asymptotic Zero-Transition Activity Encoding for Address Buses in Low-Power Microprocessor-Based Systems”, IEEE 7th Great Lakes Symposium on VLSI, Urbana, IL, pp. 77-82, Mar. 1997.

[3] E. Musoll, T. Lang, and J. Cortadella. “Working-Zone Encoding f or reducing the energy in microprocessor address buses”. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 6, no. 4, Dec 1998

[4] E. Musoll, T. Lang, and J. Cortadella. “Exploiting the locality of memory references to reduce the address bus energy”, Proceedings of ISPLED, pp. 202-207, Monterey CA, August 1997.

[5] M. Stan and W. Burleson. “Low-power encodings for global communications in CMOS VLSI”, IEEE Trans. on VLSI Systems., pages 444–455, 1997.

[6] L. Benini, G. D. Micheli, E. Macii, M. Poncino, and S. Quer. “Power optimization of core-based systems by address bus encoding”, IEEE Transactions on Very Large Scale Integration, 6(4), Dec. 1998.

[7] Tomas Akenine-Moller, Jacob Strom. “Graphics Processing Units for Handhelds”, Proceedings of the IEEE Vol. 96, No. 5, pp. 779-789 May 2008.

[8] L. Benini, G. D. Micheli, E. Macii, D. Sciuto, and C. Silvano. “Address bus encoding techniques for system-level power optimization”, In IEEE Design Automation and Test Conference in Europe, pp 861–866, Paris, Feb. 1998.

[9] Giuseppe Ascia, Vincenzo Catania, Maurizio Palesi. “A Genetic Bus Encoding Technique for Power Optimization of Embedded Systems”, J.J. Chico and E. Macii (Eds.) PATMOS 2003, LNCS 2799, pp. 21–30, 2003. Springer-Verlag Berlin Heidelberg 2003.

[10] C. Su, C. Tsui, and A. Despain. “Saving power in the control path of embedded processors”, IEEE Design and Test of computers, 11(4):24–30, 1994

[11] Wei-Chung Cheng, Massoud Pedram. “Memory Bus Encoding for Low Power: A Tutorial”. ISQED 2001: 199-204

[12] Abinesh R., Bharghava R., M.B. Srinivas, “Transition Inversion Based Low Power Data Coding Scheme for Synchronous Serial Communication”, isvlsi, pp.103-108, 2009 IEEE Computer Society Annual Symposium on VLSI, 2009

[13] Jun Yang, Rajiv Gupta, Chuanjun Zhang. “Frequent value encoding for low power data buses”. ACM Trans. Design Autom. Electr. Syst. 9(3): 354-384 (2004)

[14] Saneei M, Afzali-Kusha A, Navabi Z. “Serial Bus Encoding For Low Power Applications”, International Symposium on System-On-Chip, pp. 1-4, November 2006.

[15] Kangmin Lee, Se-Joong Lee, Hoi-Jun Yoo. “SILENT: serialized low energy transmission coding for on-chip interconnection networks”. ICCAD 2004: 448-451

AN IOT BASED SOLAR VEHICLE WITH OBSTACLE AVOIDANCE NIGHT SPEED LIMITTER AND ACCIDENT DETECTION USING GSM AND GPS

MAMIDI PRUDHVI RAJ[1]

[1]*PG Student, Dept. of ECE*, MRCE, Hyderabad, India, Email: prudhvi7269@gmail.com

ABSTRACT:

The purpose of our work is to discover the vehicle accident location through sending a message using a system which is placed within automobile system. Moreover there is a rapid rise in the event of the Roadway accident. This paper is about a system which is created to automatically detect a mishap as well as inform the nearest health centers and medical solutions about it. This system can also locate the area of the mishap so that the clinical solutions can be directed right away in the direction of it. The objective of this paper is to build up a Car unintended tracking system utilizing RESONANCE SESNOR, GPS as well as GSM Innovation. The system can be adjoined with the car alarm system and also notify the owner on his mobile phone. This discovery and also messaging system is made up of an Aurdino receiver, Microcontroller and a GSM Modem. GPS Receiver obtains the area information from satellites in the form of latitude and longitude. The GSM modem sends out an SMS to the predefined mobile number and informs about this accident. This enable it to keep an eye on the mishap circumstances and it can immediately signals the police/ambulance service with the location of mishap. The entire modules are can run with help of solar energy with battery power storage system (BESS).

Keywords: GSM, GPS, Vibration sensor, accident, latitude and longitude, location.

1. INTRODUCTION:

The usage of vehicle mobiles has boosted linearly over the past years, which boosted in the danger of human life. This is because as a result of the inadequate emergency facilities. In this paper we are using an alarm system which assists in improving the emergency situation system of the crash system. This system identifies the crash event and the collaborated of the crash are messaged to the rescue team. A changing system is used turn off in case there is no origin. The Crash is discovered with the help of RESONANCE SESNOR Sensing Unit and Vibration Sensing Unit. The Angle in which the automobile has rolled off is suggested through a message. This Application helps in supplying possible service to the bad emergency situation assists in. The function of the project is to locate the car where it is and also locate the vehicle by means of sending a message using a system which is placed inside of automobile system Most of the moments we may not have the ability to find crash area because we don't recognize where crash will happen. In order to give treatment for damaged individuals, initially we require recognizing where the crash occurred through place tracking and also sending a message to your associated one or to the emergency situation solutions. So in this job we are making use of the standard microcontroller for inexpensive as well as additionally for easy understanding. Here we made use of setting up programs for far better accuracy and GPS as well as GSM components which assists to map the car anywhere on the world. The specific area of the automobile is sent out to our remote devices (cellphones) making use of GSM modem. We are in the procedure of solving this issue by suggesting a reliable service and to lower the loss of lives as long as possible. In our concept, the layout of the system

assist us to find crashes in significantly less time and also move the basic information to the emergency treatment centre within a few secs covering the geographical works with, the moment and the angle where the vehicle had met with a mishap. This sharp message is sent to the rescue team (rescue) and the family within the brief duration. This real time application saves numerous useful lives. The message is sent with the GSM component and the area of the basic idea is to center the vehicle system by obtaining the live placement of the vehicle through GPS as well as send out the details via GSM component by means of SMS service with an added attribute of GPRS transmission to the tracking center via usage of web [M.AL-Rousan, A. R. AI-Ali as well as K. Darwish et al, 2004] Using microcontroller, this project has actually been designed. It made use of EEPROM to keep the phone numbers. Crash is discovered with the help of the GPS component. The accident can be identified precisely with the help of both Micro electro mechanical system (VIBRATION SESNOR) sensing unit and also vibration sensor. The Angle of the surrender of the cars and truck can likewise be understood by the message via the RESONANCE SESNOR sensing unit.

2. RELATED STUDY:

Currently requirements, we cannot find where the mishap has happened as well as for this reason no details pertaining to it, causing the death of an individual. The research work is taking place for tracking of the vehicle also in dark clumsy locations where there is no network for getting the signals. In literary works, a number of strategies to provide security as well as security via monitoring the lorry's real time precise positioning and also info utilizing different technology-nologies have been proposed. An excellent survey of using GPS, GSM

and GIS has actually been given in [IoanLita, Ion BogdanCioc, Daniel AlexandruVisan et alia, 2006] and also [Mrs. Ramya Kulandaivel, P.Ponmalar, B.Geetha, G.Saranya et alia, 2012] The general mechanism is to give the live geographical placement of an automobile utilizing GPS receiver and send this info to GSM facility via configurable software, this is all done by the monitoring facility which is functioning as a control unit that is attached not only by an optical wire but likewise linked wirelessly with TCP/IP protocols. The monitoring center disperses the data to the client in a reasonable format as well as it also keeps the travelling documents as well as presents the actual time information concerning automobile on digital map with GIS system [IoanLita, Ion BogdanCioc, Daniel AlexandruVisan et al, 2006] Another strategy is that automobile terminal consists of a GPS receiver which draws out details about setting with GPS satellites and also sends it via GSM network as well as to the nerve center which reads.

3. WORKING METHODOLOGY:

We propose an intelligent vehicle system for mishap avoidance and making the world a much better as well as refuge to live. Easy IR sensor is reliable for finding human or animals as well as this strategy certainly can conserve lots of lives. Pre collision detection system have to be equipped with mix of different sensors. Discovering humans or pets including challenges will certainly give us a far better service to reduce the death of human in road crash. INFLUENCE OF THE RECOMMENDED SERVICE: Presently criteria, we can not spot where the mishap has actually taken place as well as therefore no info pertaining to it, leading to the fatality of a person. The research study work is taking place for tracking the setting of the automobile even in dark

clumsy areas where there is no network for obtaining the signals. In this job GPS is used for tracking the placement of the vehicle, GSM is utilized for sending the message. Thus with this job application we can find the position of the vehicle where the accident has happened to make sure that we can provide the first aid as early as possible. This task provides lorry crash discovery and sharp system with SMS to the individual defined mobile numbers.

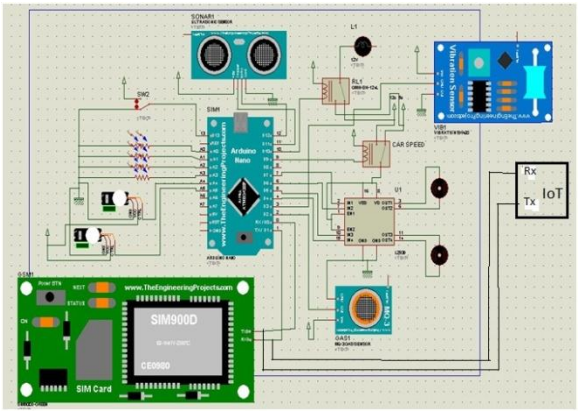


Fig.3.1. Schematic diagram of our proposed system.

When the system is turned on, LED will certainly get on suggesting that power is provided to the circuit. When the IR sensors that we are making use of in our task sense any kind of barrier, they send out disrupt to microcontroller. The GPS gets the area of the vehicle that met a crash and offers the information back. These details will be sent out to a mobile number through a message. This message will certainly be obtained utilizing GSM modem existing in the circuit. The message will certainly give the information of longitude and also latitude worth. Making use of these worth the placement of the vehicle can be estimated.

GPS LOCATION:

The Global Positioning System is a room based worldwide navigation satellite system that supplies reliable location (crash area) and also times anywhere on the world. The GPS satellites act as a referral factor from which receivers on the ground identify their setting. The essential navigating concept is based on the dimension of pseudo ranges in

between the user as well as 4 satellites.

Fig.3.2. GPS module.

ULTRASONIC SENSOR:

The Ultrasonic Sensing unit sends a high-frequency audio pulse and after that times how much time it considers the echo of the sound to mirror back. The sensor has 2 openings on its front. One opening up transmits ultrasonic waves, (like a little audio speaker), the other receives them, (like a little microphone).

Fig.3.3. Ultrasonic sensor.

ALCOHOL SENSOR:

The MQ-3 alcohol gas sensing unit includes overall 6-pins including A, H, B and the various other three pins are A, H, B out of the overall 6-pins we utilize just 4 pins. The two pins A, H are made use of for the home heating purpose and also the various other two pins are made use of for the ground and power. There is a heating unit inside the sensor, which is made up of aluminium oxide, tin dioxide. It has warmth coils to produce heat, and also hence it is used as a heat sensor. The below diagram shows the pin representation and the arrangement of the MQ-3 alcohol sensor.

Fig.3.4. Alcohol sensor.

SOLAR TRACKER:

Trackers straight photovoltaic panels or components toward the sun. These tools alter their orientation throughout the day to adhere to the sun's path to make the most of energy capture. In photovoltaic or pv systems, trackers aid lessen the angle of incidence (the angle that a beam makes with a line vertical to the surface) between the inbound light and the panel, which raises the amount of power the installation creates. Focused solar photovoltaic's and focused solar thermal have optics that straight accepts sunshine, so solar trackers need to be angled appropriately to accumulate energy.

Fig.3.4.Solar dual direction operation.

The Arduino Nano controller is used to control all the sensor modules. A dual axis tracking prototype is developed to capture the maximum sun rays by tracking the movement of the sun in four different directions. The car will move only when the person wears the seat belt and doesn't sip ant alcohol. For this, we have a seat belt and alcohol sensors. The ultrasonic sensor is used to measure the distance if any obstacle comes close to the vehicle and car will turn either left or right automatically. At the time if the vehicle goes with over speed then automatically the headlights of the vehicle goes dim.

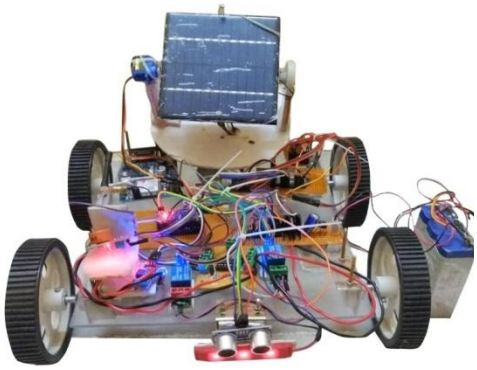


Fig.3.5. Prototype model.

If the vehicle goes in normal speed then headlights of the vehicle are blown brighter, this operation can be done by Relays and switches. If the vehicle met with any accident vibration sensor will sense it, an alert SMS will be sent to people like friends, ambulance, Police through GSM module. By using ESP-12 Wi-Fi module, and GPS module the whole information about the vehicle means its speed, condition all are updated in the Android Application. Motor drivers are used to driving the motors with the same voltage.

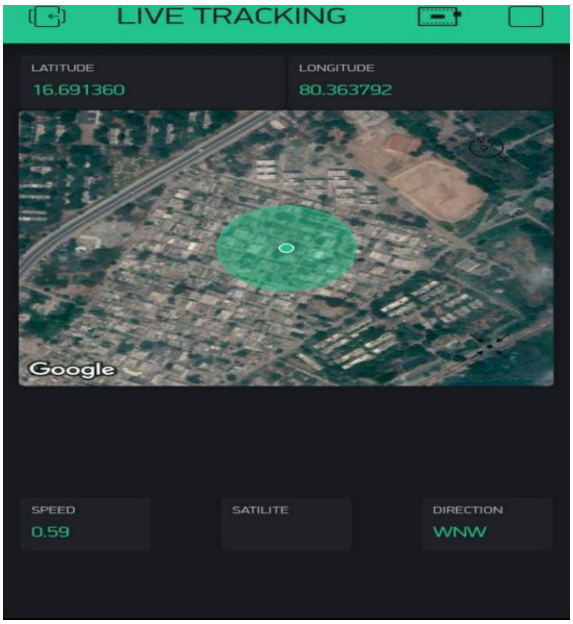


Fig.3.6. Final GPS output.

The listed below number shows that at the point when accident struck the car Vibration Sensing unit, which finds the accident and also subsequently sends out the signals to Arduino. At this point the Arduino takes control and also begins accumulating the collaborates gotten from the controller which are later sent to the Central Emergency Situation Monitoring Terminal by utilizing the GSM Component. After that the alert message will be sent on signed up mobile number through gsm.

5. CONCLUSION:

The proposed system is established to provide the information about the accident occur with location of the crash, prevents the drunk & drive accident. I conclude that my theory was correct, that the angle of the photovoltaic panel dealing with directly at the sunlight made the solar energy automobile go the fastest. The angle of the solar panel will be readjusted with respect to the activity of the sun.The car doesn't relocate up until the chauffeur wear the seat-belt and also doesn't consume any alcohol. If the chauffeur goes fast during night time the head lights of the car will immediately lower as well as get into typical circumstance if the he enters normal speed. If any other automobile comes closer to our automobile instantly an alert message will certainly be given to the driver and progressively the speed of the lorry will certainly be lowered.

REFERENCES:

[1] Vikram Singh Kushwaha, Deepa Yadav, Abuyeed Topinkatti, Amrita Kumari. "Auto Mishap Discovery System utilizing GPS And GSM", Volume 2, Concern 1(Jan-Feb 2015), PP12-17

[2] Nimisha Chaturvedi, Pallika Srivastava. "Automatic Lorry Accident Detection and Messaging System Making Use Of GSM as well as GPS Modem ", Volume: 05 Issue: 03|Mar-2018

[3] C.Prabha, R.Sunitha, R.Anitha. "Automatic Car Mishap Discovery and also Messaging System Utilizing GSM and GPS Modem", Vol. 3, Concern 7, July 2014

[4] Hoang Dat Pham, MichealDrieberg, Chi Cuong Nguyen, "Advancement of car tracking system making use of GENERAL PRACTITIONER and GSM modem ", Seminar: 2013 IEEE Conference on Open Equipment (ICOS).

[5] Lih-Jen Kau, Member, IEEE, and also Chih-Sheng Chen, "A Smart Phone-Based Pockert Loss Crash Discovery, Positioning As Well As Rescue System", Dec 2013.

[6] G. Acampora, D. J. Cook, P. Rashidi, A. V. Vasilakos, "A Study on Ambient Intelligence in Health Care", Proceedings of the IEEE, pp. 2470-2494, Vol. 101, No. 12, Dec. 2013.

[7] Malik Tubaishat, Qi Qi, Yi Shang, Hongchi Shi "Wireless Sensor-Based Traffic Signal Control" IEEE CCNC 2008 process 1-4244-1457-1/ 08.

[8] "Automatic crash notice system using gsm and also gps modems with 3g technology for video monitoring" International Journal of Emerging Fads in Electric and also Electronics (IJETEE) Vol. 1, Issue. 2, March-2013.

[9] E.Krishna Priya, P.Manju, V.Mythra, "IoT Based Lorry Tracking and Accident Discovery System" International journal of Innovative Study in Computer as well as Communication Engineering, (An ISO 3297:2007 Licensed association) Vol.5, Concern 3.

[10] Sri Krishna Chaitanya Varma, Poornesh, Tarun Varma, Harsha; Automatic Car Accident Detection And Also Messaging System Utilizing GPS and also GSM Modems; International Journal of Scientific & Design Research study, Volume 4, Problem 8, August-2013.

DLAU: A Scalable Deep Learning Accelerator Unit on FPGA

VAJJA PARMESH,PG student,Dept of ECE,MRCE,Hyderabad,mail:-vajjaparamesh12@gmail.com

Abstract—As the emerging field of machine learning, deep learning shows excellent ability in solving complex learning problems. However, the size of the networks becomes increasingly large scale due to the demands of the practical applications, which poses significant challenge to construct a high performance implementations of deep learning neural networks. In order to improve the performance as well to maintain the low power cost, in this paper we design DLAU, which is a scalable accelerator architecture for large-scale deep learning networks using FPGA as the hardware prototype. The DLAU accelerator employs three pipelined processing units to improve the throughput and utilizes tile techniques to explore locality for deep learning applications. Experimental results on the state-of-the-art Xilinx FPGA board demonstrate that the DLAU accelerator is able to achieve up to 36.1x speedup comparing to the Intel Core2 processors, with the power consumption at 234mW.

Index Terms—FPGA; Deep Learning; neural network; hard-ware accelerator.

I. INTRODUCTION

IN the past few years, machine learning has become perva-sive in various research fields and commercial applications, and achieved satisfactory products. The emergence of deep learning speeded up the development of machine learning and artificial intelligence. Consequently, deep learning has become a research hot spot in research organizations [1].

In general, deep learning uses a multi-layer neural network model to extract high-level features which are a combination of low-level abstractions to find the distributed data features, in order to solve complex problems in machine learning. Currently the most widely used neural models of deep learning are Deep Neural Networks (DNNs) [2] and Convolution Neural Net-works (CNNs) [3], which have been proved to have excellent capability in solving picture recognition, voice recognition and
other complex machine learning tasks.

However, with the increasing accuracy requirements and complexity for the practical applications, the size of the neural networks becomes explosively large scale, such as the Baidu Brain with 100 Billion neuronal connections, and the Google cat-recognizing system with 1 Billion neuronal connections. The explosive volume of data makes the data centers quite power consuming. In particular, the electricity consumption of data centers in U.S. are projected to increase to roughly 140 billion kilowatt-hours annually by 2020 [4]. Therefore, it poses significant challenges to implement high performance deep learning networks with low power cost, especially for large-scale deep learning neural network models. So far, the state-of-the-art means for accelerating deep learning algorithms are Field-Programmable Gate Array (FPGA), Application Spe-cific Integrated Circuit (ASIC), and Graphic Processing Unit (GPU). Compared with GPU acceleration, hardware accel-erators like FPGA and ASIC can achieve at least moderate performance with lower power consumption. However, both FPGA and ASIC have relatively limited computing resources, memory, and I/O bandwidths, therefore it is challenging to develop complex and massive deep neural networks using hardware accelerators. For ASIC, it has a longer development cycle and the flexibility is not satisfying. Chen et al presents a ubiquitous machine-learning hardware accelerator called DianNao [6], which initiated the field of deep learning pro-cessor. It opens a new paradigm to machine learning hardware accelerators focusing on neural networks. But DianNao is not implemented using reconfigurable hardware like FPGA, therefore it cannot adapt to different application demands. Currently around FPGA acceleration researches, Ly and Chow
[5] designed FPGA based solutions to accelerate the Re-stricted Boltzmann Machine (RBM). They created dedicated hardware processing cores which are optimized for the RBM algorithm. Similarly Kim et al [7] also developed a FPGA based accelerator for the restricted Boltzmann machine. They use multiple RBM processing modules in parallel, with each module responsible for a relatively small number of nodes. Other similar works also present FPGA based neural network accelerators [9]. Qi et al. present a FPGA based accelerator [8], but it cannot accommodate changing network size and network topologies. To sum up, these studies focus on implementing a particular deep learning algorithm efficiently, but how to increase the size of the neural networks with scalable and flexible hardware architecture has not been properly solved.

To tackle these problems, we present a scalable deep learning accelerator unit named DLAU to speed up the kernel computational parts of deep learning algorithms. In particular, we utilize the tile techniques, FIFO buffers, and pipelines to minimize memory transfer operations, and reuse the comput-ing units to implement the large-size neural networks. This approach distinguishes itself from previous literatures with following contributions:

1. In order to explore the locality of the deep learning

TABLE I
PROFILING OF HOT SPOTS OF DNN

Algorithms	Matrix Multiplication	Activation	Vector
Feedforward	98.60%	1.40%	
RBM	98.20%	1.48%	0.30%
BP	99.10%	0.42%	0.48%

application, we employ tile techniques to partition the large scale input data. The DLAU architecture can be configured to operate different sizes of tile data to leverage the trade-offs between speedup and hardware costs. Consequently the FPGA based accelerator is more scalable to accommodate different machine learning applications.

2. The DLAU accelerator is composed of three fully pipelined processing units, including TMMU, PSAU, and AFAU. Different network topologies such as CNN, DNN, or even emerging neural networks can be composed from these basic modules. Consequently the scalability of FPGA based accelerator is higher than ASIC based accelerator.

II. TILE TECHNIQUES AND HOT SPOT PROFILING

Restricted Boltzmann Machines (RBMs) have been widely used to efficiently train each layer of a deep network. Normally a deep neural network is composed of one input layer, several hidden layers and one classifier layer. The units in adja-cent layers are all-to-all weighted connected. The prediction process contains feedforward computation from given input neurons to the output neurons with the current network config-urations. Training process includes pre-training which locally tune the connection weights between the units in adjacent layers, and global training which globally tune the connection weights with Back Propagation process.

The large-scale deep neural networks include iterative com-putations which have few conditional branch operations, there-fore they are suitable for parallel optimization in hardware. In this paper we first explore the hot spot using the profiler. Results in Fig. I illustrates the percentage of running time including Matrix Multiplication (MM), Activation, and Vector operations. For the representative three key operations: feed forward, Restricted Boltzmann Machine (RBM), and back propagation (BP), matrix multiplication play a significant role of the overall execution. In particular, it takes 98.6%, 98.2%, and 99.1% of the feed forward, RBM, and BP operations. In comparison, the activation function only takes 1.40%, 1.48%, and 0.42% of the three operations. Experimental results on profiling demonstrate that the design and implementation of MM accelerators is able to improve the overall speedup of the system significantly.

However, considerable memory bandwidth and computing resources are needed to support the parallel processing, con-sequently it poses a significant challenge to FPGA implemen-tations compared with GPU and CPU optimization measures. In order to tackle the problem, in this paper we employ tile techniques to partition the massive input data set into tiled subsets. Each designed hardware accelerator is able to buffer the tiled subset of data for processing. In order to support the large-scale neural networks, the accelerator architecture are reused. Moreover, the data access for each tiled subset can run in parallel to the computation of the hardware accelerators.

Algorithm 1 Pseudocode Code of the Tiled

```
Inputs Require:
  Ni: the number of the input neurons
  No: the number of the output neurons
  Tile_Size: the tile size of the input data
  batchsize: the batch size of the input
  data for n = 0; n < batchsize; n + + do
    for k = 0; k < Ni; k+ = T ile Size do
      for j = 0; j < No; j + + do
        y[n][j] = 0;
        for i = k; i < k + T ile Size&&i < Ni; i + + do
          y[n][j]+ = w[i][j] x[n][i]
          if i == Ni       1 then
            y[n][j] = f(y[n][j]);
          end if
        end for
      end for
    end for
  end for
```

In particular, for each iteration, output neurons are reused as the input neurons in next iteration. To generate the output neurons for each iteration, we need to multiply the input neurons by each column in weights matrix. As illustrated in Algorithm 1, the input data are partitioned into tiles and then multiplied by the corresponding weights. Thereafter the calculated part sum are accumulated to get the result. Besides the input/output neurons, we also divided the weight matrix into tiles corresponding to the tile size. As a consequence, the hardware cost of the accelerator only depends on the tile size, which saves significant number of hardware resources. The tiled technique is able to solve the problem by imple-menting large networks with limited hardware. Moreover, the pipelined hardware implementation is another advantage of FPGA technology compared to GPU architecture, which uses massive parallel SIMD architectures to improve the overall performance and throughput. According to the profiling results depicted in Table I, during the prediction process and the training process in deep learning algorithms, the common but important computational parts are matrix multiplication and activation functions, consequently in this paper we implement the specialized accelerator to speed up the matrix multiplica-tion and activation functions.

III. DLAU ARCHITECTURE AND EXECUTION MODEL

Fig. 1 describes the DLAU system architecture which contains an embedded processor, a DDR3 memory controller, a DMA module, and the DLAU accelerator. The embedded processor is responsible for providing programming interface to the users and communicating with DLAU via JTAG-UART. In particular it transfers the input data and the weight matrix to internal BRAM blocks, activates the DLAU accelerator, and returns the results to the user after execution. The DLAU is integrated as a standalone unit which is flexible and adaptive

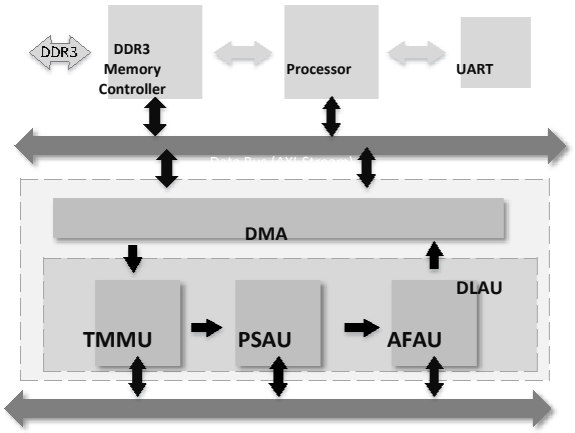


Fig. 1. DLAU Accelerator Architecture.

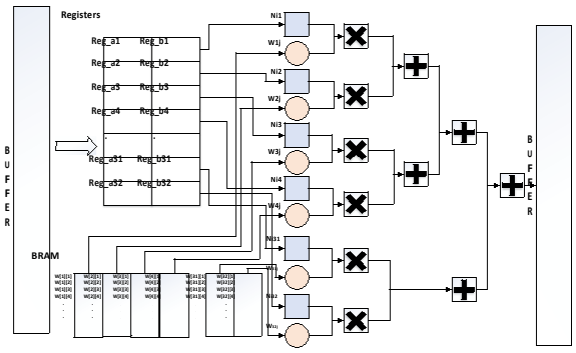


Fig. 2. TMMU Schematic Diagram.

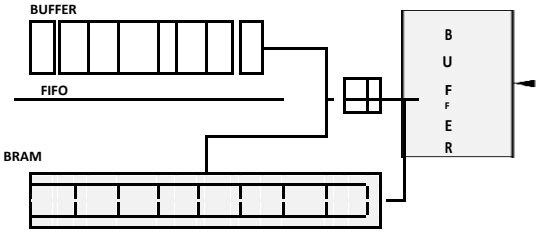


Fig. 3. PSAU Schematic Diagram

to accommodate different applications with configurations. The DLAU consists of 3 processing units organized in a pipeline manner: Tiled Matrix Multiplication Unit (TMMU), Part Sum Accumula2tion Unit (PSAU), and Activation Function Acceleration Unit (AFAU). For execution, DLAU reads the tiled data from the memory by DMA, computes with all the three processing units in turn, and then writes the results back to the memory.

In particular, the DLAU accelerator architecture has follow-ing key features:

FIFO Buffer: Each processing unit in DLAU has an input buffer and an output buffer to receive or send the data in FIFO. These buffers are employed to prevent the data loss caused by the inconsistent throughput between each processing unit.

Tiled Techniques: Different machine learning applications may require specific neural net-work sizes. The tile technique is employed to divide the large volume of data into small tiles that can be cached on chip, therefore the accelerator can be adopted to different neural network size. Consequently the FPGA based accelerator is more scalable to accommodate different machine learning applications.

Pipeline Accelerator: We use stream-like data passing mechanism (e.g. AXI-Stream for demonstration) to transfer data between the adjacent processing units, therefore TMMU, PSAU, and AFAU can compute in streaming-like manner. Of these three computational modules, TMMU is the primary computational unit, which reads the total weights and tiled nodes data through DMA, performs the calculations, and then transfers the intermediate Part Sum results to PSAU. PSAU collects Part Sums and performs accumulation. When the accumulation is completed, results will be passed to AFAU. AFAU performs the activation function using piecewise linear interpolation methods. In the rest of this section, we will detail the implementation of these three processing units respectively.

A. TMMU architecture

Tiled Matrix Multiplication Unit (TMMU) is in charge of multiplication and accumulation operations. TMMU is spe-cially designed to exploit the data locality of the weights and is responsible for calculating the Part Sums. TMMU employs an input FIFO buffer which receives the data transferred from DMA and an output FIFO buffer to send Part Sums to PSAU. Fig. 2 illustrates the TMMU schematic diagram, in which we set tile size=32 as an example. TMMU firstly reads the weight matrix data from input buffer into different BRAMs in 32 by the row number of the weight matrix (n=i%32where n refers to the number of BRAM, and i is the row number of weight matrix). Then, TMMU begins to buffer the tiled node data. In the first time, TMMU reads the tiled 32 values to registers Reg a and starts execution. In parallel to the computation at every cycle, TMMU reads the next node from input buffer and saves to the registers Reg b. Consequently the registers Reg a and Reg_b can be used alternately.

For the calculation, we use pipelined binary adder tree structure to optimize the performance. As depicted in Fig. 2, the weight data and the node data are saved in BRAMs and registers. The pipeline takes advantage of time-sharing the coarse-grained accelerators. As a consequence, this im-plementation enables the TMMU unit to produce a Part Sum result every clock cycle.

B. PSAU architecture

Part Sum Accumulation Unit (PSAU) is responsible for the accumulation operation. Fig. 3 presents the PSAU architecture, which accumulates the part sum produced by TMMU. If the Part Sum is the final result, PSAU will write the value to output buffer and send results to AFAU in a pipeline manner. PSAU can accumulate one Part Sum every clock cycle, therefore the throughput of PSAU accumulation matches the generation of the Part Sum in TMMU.

C. AFAU architecture

Finally, Activation Function Acceleration Unit (AFAU) implements the activation function using piecewise linear interpolation (y=ai*x+bi, x2[x_1,x_{i+1})). This method has been widely applied to implement activation functions with negligible accuracy loss when the interval between x_i and x_{i+1} is insignificant. Eq. (1) shows the implementation of sigmoid function. For x>8 and x -8, the results are sufficiently close to the bounds of 1 and 0, respectively. For the cases in -8<x 0 and 0<x 8, different functions are configured. In total we divide the sigmoid function into four segments.

$$f(x) = \begin{cases} 1 + a[b^{x}_{k} c_{x}]x \quad b[b^{x}_{kx} \quad c] & \text{if } 8 < x \quad 0 \\ > \quad 0 & \text{if } x \quad 8 \\ < \quad b\_c \quad b\_ \ c_{1} & \text{if } x > 8 \\ : \end{cases} \qquad (1)$$

Similar to PSAU, AFAU also has both input buffer and output buffer to maintain the throughput with other processing units. In particular, we use two separate BRAMs to store the values of a and b. The computation of AFAU is pipelined to operate sigmoid function every clock cycle. As a consequence, all the three processing units are fully pipelined to ensure the peak throughput of the DLAU accelerator architecture.

IV. Experiments and Data Analysis

In order to evaluate the performance and cost of the DLAU accelerator, we have implemented the hardware prototype on the Xilinx Zynq Zedboard development board, which equips ARM Cortex-A9 processors clocked at 667MHz and pro-grammable fabrics. For benchmarks, we use the Mnist data set to train the 784 M N 10 Deep Neural Networks in Matlab, and use M N layers weights and nodes value for the input data of DLAU. For comparison, we use Intel Core2 processor clocked at 2.3GHz as the baseline.

In the experiment we use Tile size=32 considering the hardware resources integrated in the Zedboard development board. The DLAU computes 32 hardware neurons with 32 weights every cycle. The clock of DLAU is 200MHz (one cycle takes 5ns). Three network sizes—64 64, 128 128, and 256 256 are tested.

A. Speedup Analysis

We present the speedup of DLAU and some other similar implementations of the deep learning algorithms in Table II. Experimental results demonstrate that the DLAU is able to achieve up to 36.1x speedup at 256 256 network size. In comparison, Ly&Chows work [5] and Kim et.als work [7] present the work only on Restricted Boltzmann Machine algorithms, while the DLAU is much more scalable and flexible. DianNao [6] reaches up to 117.87x speedup due to its high working frequency at 0.98GHz. Moreover, as DianNao is hardwired instead of implemented on a FPGA platform, therefore it cannot efficiently adapt to different neural network sizes.

Fig. 4 illustrates the speedup of DLAU at different network sizes-64 64, 128 128, and 256 256 respectively. Experi-mental results demonstrate a reasonable ascendant speedup

TABLE II
Comparisons between Similar Approaches

Work	Network	Clock	Speedup	Baseline
Ly&Chow [5]	256 256	100MHz	32	2.8GHz P4
Kim et.al [7]	256 256	200MHz	25	2.4GHz Core2
DianNao [6]	General	0.98GHz	117.87	2GHz SIMD
Zhang et.al [3]	256 256	100MHz	17.42	2.2GHz Xeon
DLAU	256 256	200MHz	36.1	2.3GHz Core2

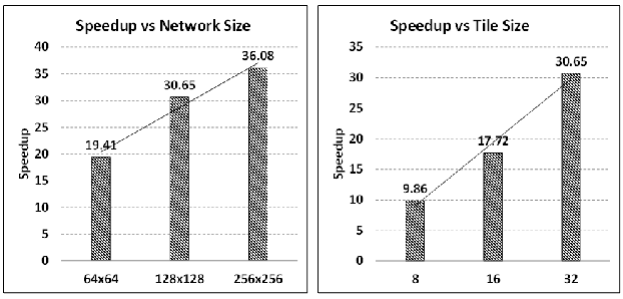


Fig. 4. Speedup at Different Network Sizes and Tile Sizes.

TABLE III
Resource Utilization of DLAU at 32 32 Tile Size

Component	BRAMs	DSPs	FFs	LUTs
TMMU	32	158	25356	32461
PSAU	1	2	754	632
AFAU	2	7	2216	3291
Total	35	167	28326	36384
Available	280	220	106400	53200
Utilization	12.5%	75.9%	26.6%	68.4%

with the growth of neural networks sizes. In particular, the speedup increases from 19.2x in 64 64 network size to 36.1x at the 256 256 network size. The right part of Fig. 4 illustrates how the tile size has an impact on the performance of the DLAU. It can be acknowledged that bigger tile size means more number of neurons to be computed concurrently. At the network size of 128 128, the speedup is 9.2x when the tile size is 8. When the tile size increases to 32, the speedup reaches 30.5x. Experimental results demonstrate that the DLAU framework is configurable and scalable with dif-ferent tile sizes. The speedup can be leveraged with hardware cost to achieve satisfying trade-offs.

B. Resource utilization and Power

Table III summarizes the resource utilization of DLAU in 32 32 tile size including the BRAM resources, DSPs, FFs, and LUTs. TMMU is much more complex than the rest two hardware modules therefore it consumes most hardware resources. Taking the limited number of hardware logic resources provided by Xilinx XC7Z020 FPGA chip, the overall utilization is reasonable. The DLAU utilizes 167 DSP blocks due to the use of the Floating-point addition and the Floating-point multiplication operations.

Table IV compares the resource utilization of DLAU with other two FPGA based literatures. Experimental results depict

TABLE IV
RESOURCE COMPARISONS BETWEEN SIMILAR APPROACHES

Implementation	FPGA	BRAMs	DSPs	FFs	LUTs
Ly&Chow [5]	XC2VP70	257	N/A	30403	29885
Kim et.al [7]	N/A	589824	18	11790	7662
DLAU	XC7Z020	35	167	28326	36384

TABLE V
POWER CONSUMPTION OF THE UNITS

Component	Power	Component	Power
Accelerator-TMMU	189mW	Processor	1307mW
Accelerator-PSAU	5mW	DDR Controller	177mW
Accelerator-AFAU	25mW	Peripherals	26mW
Accelerator-DMA	15mW	Clocks	70mW
Accelerator-Total	234mW	System Total	1814mW

that our DLAU accelerator occupies similar number of FFs and LUTs to Ly&Chow's work [5], while it only consumes 35/257=13.6% on the BRAMs. Comparing to the Kim et.al's work [7], the BRAM utilization of DLAU is insignificant. This is due to the tile techniques so that large scale neural networks can be divided into small tiles, therefore the scalability and flexibility of the architecture is significantly improved.

In order to evaluate the power consumption of accelerator, we use Xilinx Vivado tool set to achieve power cost of each processing unit in DLAU and the DMA module. The results in Table IV-B depict that the total power of DLAU is only 234mW, which is much lower than that of DianNao (485mW). The results demonstrate that the DLAU is quite energy efficient as well as highly scalable compared to other accelerating techniques. To compare the energy and power between FPGA based accelerator and GPU based accelera-tors, we also implement a prototype using the state-of-the-art NVIDIA Tesla K40c as the baseline. K40c has 2880 stream cores working at peak frequency 875MHz, and the Max Memory Bandwidth is 288 (GB/sec). In comparison, we only employ 1 DLAU on the FPGA board working at 100MHz. In order to evaluate the speedup of the accelerators in a real deep learning applications, we use DNN to model 3 benchmarks, including Caltech101, Cifar-10, and MNIST, respectively. Fig. 5 illustrates the comparison between FPGA based GPU+cuBLAS implementations. It reveals that the power consumption of GPU based accelerator is 364 times higher than FPGA based accelerators. Regarding the total energy consumption, the FPGA based accelerator is 10x more energy efficient than GPU, and 4.2x than GPU+cuBLAS optimizations.

Finally Fig. 6 illustrates the floor plan of the FPGA chip. The left corner depicts the ARM processor which is hard-wired in the FPGA chip. Other modules, including different components of the DLAU accelerator, the DMA, and memory interconnect, are presented in different colors. Regarding the programming logic devices, TMMU takes most of the areas as it utilizes a significant number of LUTs and FFs.

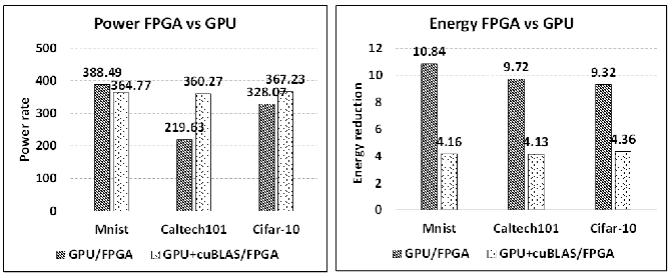


Fig. 5. Power and Energy Comparison between FPGA and GPU

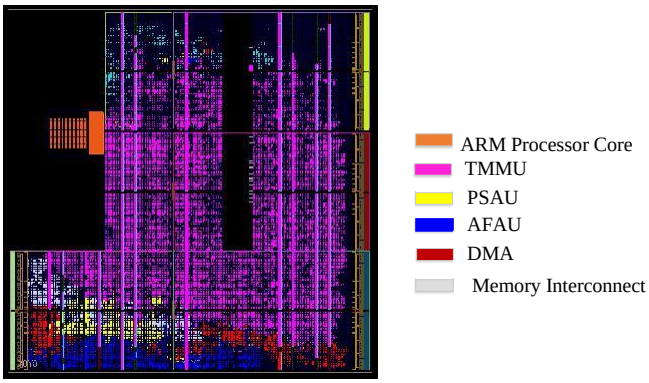


Fig. 6. Floorplan of the FPGA Chip

V. CONCLUSION AND FUTURE WORK

In this article we have presented DLAU, which is a scalable and flexible deep learning accelerator based on FPGA. The DLAU includes three pipelined processing units, which can be reused for large scale neural networks. DLAU uses tile techniques to partition the input node data into smaller sets and compute repeatedly by time-sharing the arithmetic logic. Experimental results on Xilinx FPGA prototype show that DLAU can achieve 36.1x speedup with reasonable hardware cost and low power utilization.

The results are promising but there are still some future directions, including optimization of the weight matrix and memory access. Also the trade-off analysis between FPGA and GPU accelerators is another promising direction for large scale neural networks accelerations.

REFERENCES

[1] LeCun, Y., Y. Bengio, and G. Hinton, Deep learning. Nature, 2015. 521: p. 436-444.
[2] Hauswald, J., et al. DjiNN and Tonic: DNN as a service and its implications for future warehouse scale computers. in ISCA 2015.
[3] Zhang, C., et al. Optimizing FPGA-based Accelerator Design for Deep Convolutional Neural Networks. in FPGA 2015.
[4] Thibodeau, P. Data centers are the new polluters. 2014 [cited 2015.
[5] Ly, D.L. and P. Chow, A high-performance FPGA architecture for restricted boltzmann machines, in FPGA 2009.
[6] Chen, T., et al., DianNao: a small-footprint high-throughput accelerator for ubiquitous machine-learning, in ASPLOS 2014. p. 269-284.
[7] Kim, S.K., et al. A highly scalable restricted boltzmann machine FPGA implementation. in FPL 2009.
[8] Qi Yu, et al. A Deep Learning Prediction Process Accelerator Based FPGA. CCGRID 2015: 1159-1162
[9] Jiantao Qiu, et al. Going Deeper with Embedded FPGA Platform for Convolutional Neural Network. FPGA 2016: 26-35

Division Circuit Using Reversible Logic Gates

RATHNAVATH PRIYADARSHINI
M.tech scholar , Dept of ECE,
Hyderabad,Priyadarshini.r73@gmail.com.

Abstract—In the recent years, reversible approachis becoming widely used in many domains, such as quantum computing, optical computing and ultra-low power VLSI circuit. Division has its application in the design of reversible Arithmetic Logic Unit (ALU). In this paper, we have exhibited a novel design of division sequential circuit using reversible logic gates. The proposed design of division block is based on reversible gates with reduction of garbage outputs, constant inputs, quantum cost and hardware complexity. The comparative results demonstrate that the proposed solution have less performanceand significantly better scalability than the currentdesigns

Keywords— Low power; reversible ALU; reversible gates; quantum cost; garbage outputs; hardware complexity; constant inputs; reversible division

I. INTRODUCTION

Power consumption is an essential issue in modern day VLSI (Very-Large-Scale Integration) designs. Consumption has become the essential limit of the increasing number of transistors per chip. In the recent years, many techniques have been developed at various levels in CMOS VLSI design.

- At system level and algorithmic level, techniques such as use of dynamic voltage Scaling (DVS) [1] and dynamic power management (DPM), as well as alternate encoding methods.[2].
- At architecture level, techniques such as use of parallel structures, pipelining [3], the clock gating [4].
- At circuit level, techniques such as use of the glitch [5].
- At the technology level, techniques [6] such as VT reduction, multi-threshold voltages have beenused.

Front of these diversity methods that are dedicated to improving the power VLSI design. In this paper we have encountered the reversible logic method [7]. It has wide applications in several technologies such as the nanotechnology, low power CMOS, optical information processing, bio information and DNA computing. In 1961 Rolf Landauer [8] showed that the irreversible logic gatesor the conventional logic gates, set-to-one like AND, OR, XOR and others dissipate a certain amount of energy caused the loss for each bit information during computation (at least kTln2 for each bit of information lost, where K is the Boltzmann's constant and T is the operating temperature).

In 1973, Bennett showed that [9] in order to avoid KT*ln2 joules of energy dissipation in a irreversible circuit, it must be built using reversible logic gates only, since there is no information loss happens in reversible circuits. In fact, division module is very important in the digital processing. Division process is one of the most difficult operations in the ALU and plays a big role in digital signal processing systems. This paper focuses on the design of division circuit sequential using reversible logic gates for positiveintegers.

II. BACKGROUND

In this section, we have exhibited the essential principles definitions to reversible logic along with the analysis of quantum cost.

A. ReversibleFunctions

Reversible function can be realized by reversible logic, is a bijection. Then there is a unique one-to-one mapping between an n-input vector and a corresponding n-output vector. Avoid leading output signals of gates to more than one input (Fan_out).Agatewithkinputsandkoutputsiscalledk*k gate. An k*k reversible logic gate can be denoted as $I_K - O_K$, where, $I_K = (I_1, I_2, \ldots, I_K)$ is the input vector and $O_K = (O_1, O_2, \ldots, O_K)$ is the output vector as show in fig.1 [10]. Reversible logic circuits have theoretically zero internal power dissipation because they do not lose data.

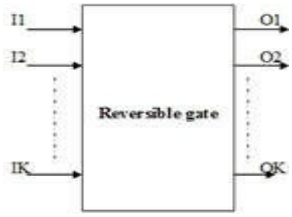


Fig.1. Reversible logic gate

B. Garbageoutputs

That defines the outputs of the reversible gate that is not used for further computations [11] [12]. The unutilized outputs from a gate are called ”garbage output” as show infig.2.

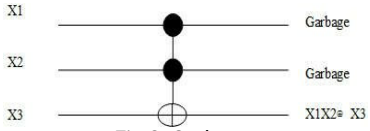


Fig.2. Garbage output

C. Constant inputs

Constant inputs (CIs) refer to the inputs which are used as control inputs and connected to logical '0' or '1' in order to obtain the logical function.
[13].

D. Quantum Cost

Quantum cost is the number of elementary 1*1 like NOT and 2*2 quantum logic gateslikeControlled ,Controlled and CNOT gates[14] [15] needed to realize the circuit.

E. Hardwarecomplexity

This refers to the total number of the logical calculations in a circuit that is measured by counting the number of AND operations, number of NOT operations and number of EX_OR operations [16]. To compute the hardware complexity of the reversible circuits we assume that:
a = Number of EX-OR gates. þ = Number of ANDgates.
6 = Number of NOTgates.

F. Synthesis of reversiblelogic

The design of reversible circuits significantly differs from the design of classical circuits. A reversible circuit should be designed using minimum number of reversible gates. One key requirement to achieve optimization is that the designed circuit must generate minimum number of garbage outputs. Equally they must use minimum number of constant inputs[17].
The fundamental rules for efficient reversible logic synthesis are:

- Reduction the number of quantum cost, garbage outputs, constant inputs and hardwarecomplexity.
- Avoid leading output signals of gates to more than one input (fan_out is notpermitted).
- Use a less number of reversible gates as conceivable to attain thegoal.

III. BASIC REVERSIBLEGATES

Various reversible gates have been previously proposed by researchers/designers till now [24-30]. Each gate has a cost associated with it called the quantum cost. The NOT gate is a 1-q-bit gates and it has a quantum cost of zero. The N-bit Controlled-Gate has quantum cost of n-1. The Feynman gate can be operates as a controlled NOT (CNOT). If A is set to ‘1’ then the gate behaves as a Not gate, else a buffer gate. Feynman gate is widely used to surmount the fan-out problem as fan-out is not allowed in the reversible logic. It has a quantum cost of 1. The quantum cost of a Double Feynman gate is 2. The quantum cost of a Toffoli gate, TR gate and Peres gate are 4. Some examples of reversible logic gates are given by Table.1 [16-20]. These reversible gates help researchers/designers to design higher complex computing circuits. In this manuscript, we employ the reversible approach to realize a Division module.

TABLE1. Examples of some reversible logic gates

Gate	Quantum cost
Feynman gate	1
Not gate	1
Tr gate	4
Toffoli gate	5
Peres gate	4
Fredkin gate	5
BHA gate	4
BHC gate	5

IV. Relatedwork

A. Division operation

Essentially, the parameters of the division operation are dividend X and divisor Y as an input, and the quotient Q and remainder R as output, With X= Q*Y+R. For division, we use shift/adder division algorithm. At every step, we Shifted the register A and X (dividend) of 1 bit to the left. If the content of the register S.A is negative, then we add the contents of register A to Y. Else we subtract the contents of register A to
Y. If the result of subtraction/addition is positive (or zero) then, the quotient bit qi = 1. Else the quotient bit qi = 0. This process is repeated ntimes.

B. Components of division circuit

- Input reversible 8 bitMUX

Figure 3 shows the 2 inputs n-bit reversible MUX where Sel is the select input, and aretwo inputs. If Sel = 0, then or if S =1, then . This

reversible MUX consists of n the applicable criteria that follow.BHA[16]gateswhichisgeneratengarbageoutputsand needs 4n quantumcost

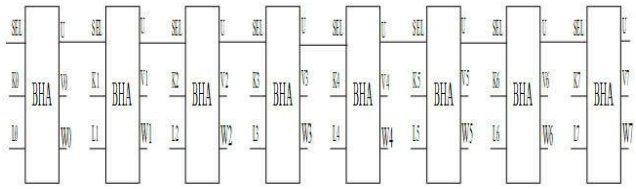
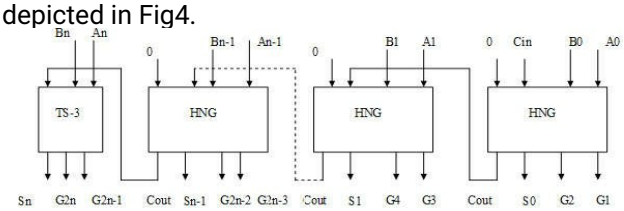


Fig 3. 2-input 8-bit reversible multiplexer circuit

- (8+1) BIT PARALLEL ADDER/SUBTRACTOR

To perform reversible realization of (n+1) adder/subtractor circuit we may use different combinations of any of the reversible logic gates. In [20] n bit reversible adder/ Subtractor using HNG gates is presented. The carry-out of the adder/ Subtractor is ignored in the proposed division circuit. Then, the implementation of (n+1) bit parallel adder/ subtractor requires n full Adder/Subtractor units and one TS-3 gate as depicted in Fig4.



II.

FIG 4. REVERSIBLE (N + 1) BIT PARALLEL ADDER WITH IGNORING OUTPUT CARRY

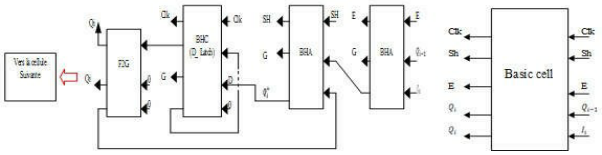
- N BIT REVERSIBLE PIPO LEFT-SHIFTREGISTERS

In n bit PIPO shift register as show in fig.15, during every clock pulse all information bits are loaded into the register. After shift operation all information bits are transferred together to their respective outputs by the same clock pulse. In [10], an n bits reversible PIPO right-shift register has been proposed. The researchers make it compatible for implementing left-shift register. Then, the basic elements used to design left shift register are multiplexer and D-latch. Fig 5 shows the proposed left shift register.

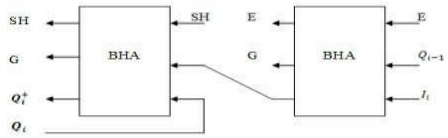
TABLE 2. Function table for reversible PIPO shift register.

SH	E	Finaloutput Q_i^+
0	0	Q_{i-1}(Leftshift)
0	1	I_i(Parallelload)
1		Q_i(Nochange)

- Basic cell for 1 bit reversible PIPO(Parallel Input-ParallelOutput)



- 3 to 1 reversibleMUX



- Basic cell for n bit reversible PIPO

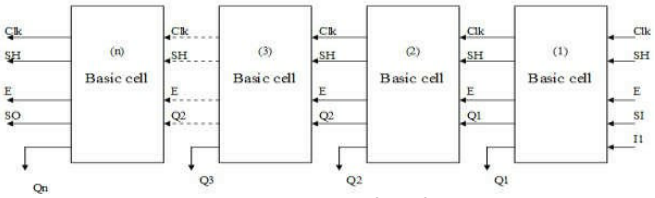


Fig 5. Reversible PIPO-left-shift register

- n- bit reversibleregister

The reversible D-Latch is used in [16] using BHC gate. Then, in order to implement n-bits reversible register we can use n BHC gates. Fig 6 shows the n-bit reversible register. It consists of n BHC gates, produces n garbage outputs, n constant inputs and needs 5n quantum cost. The hardware complexity is 4a+6þ+46.

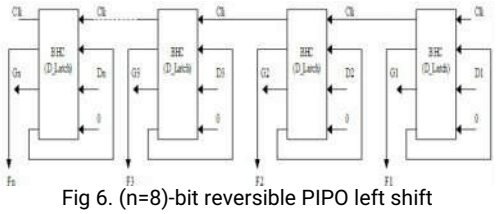


Fig 6. (n=8)-bit reversible PIPO left shift register.

- REVERSIBLE COUNTER (MOD-8)

A mod-8 counter reversible stores an integer value, and increments that value on each clock tick, and wraps around to 0 if the previous stored value was 7. The BHC gate [16] can be used as a D_ latch,where

$$Q_{(t)} = Clk.D + \overline{Clk}.Q_{(t-1)} \quad (1)$$

So we need three flip-flops D_

latch(BHC)where:D_0=1

$D_1 = Q_0$

$D_2 = Q_0.Q_1$

This reversible circuit produces a total number of 9 garbage outputs and 8 Constant inputs as show in Fig 7. It has a quantum cost of 28.

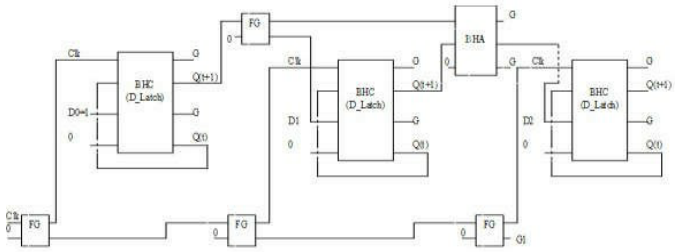


Fig 7. Reversible counter (mod-8)

- 4 Bit reversiblecomparator

In this section, two 4 bit numbers are compared with each other and the result shows that if one number is superior or fewer than other or if the two numbers are equal with each other. We use NOT gate, TR gate and BJN gate [17]. For example, assume A=A3 A2 A1 A0 and B=B3 B2 B1 B0, for comparing these two numbers, we use these finite state machine as show in Fig8.

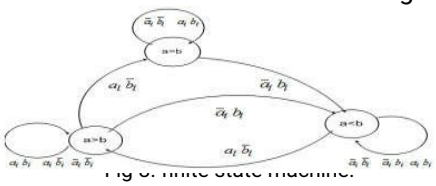


Fig 8. Finite state machine.

Reversible comparator is demonstrated in Fig 9. This circuit produces a total number of 10 constant inputs, 15 garbage outputs, and 18 gates. The quantum cost of this comparator is 38.

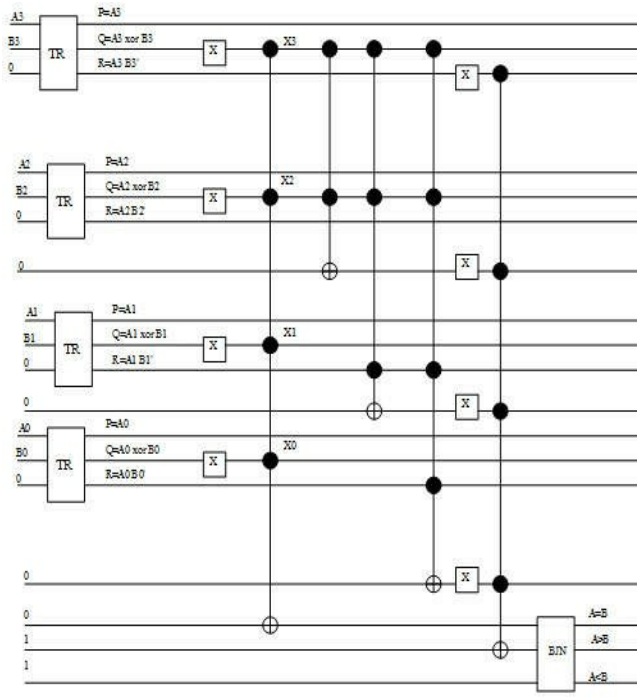


Fig 9. 4 bit reversible comparator.

Reversible n bitdivision

The reversible components are used to implement the division operation. In this paper an optimized reversible non restoring division using reversible logic gates is presented. This circuit includes (n=8).

- One n bit reversible PIPO left shiftregister
- One (n+1) bit reversible PIPO left shiftregister
- One 2 input reversible n bitMUX
- One 2 input reversible (n+1) bitMUX
- One (n+1) bit reversible parallel adder/Subtractor.

Algorithm:

Inputs:

S.A($S, A_{n-1}\ A_{n-2}\ldots\ A_0$) = 0 and
X($X_{n-1}\ X_{n-2}\ldots X_0$) =dividendand
Y($Y_{n-1}\ Y_{n-2}\ldots Y_0$)=divisor

Outputs:

Q($Q_{n-1}\ Q_{n-2}\ldots Q_0$)=quotientand
A($A_{n-1}\ A_{n-2}\ldots A_0$)=remainder.

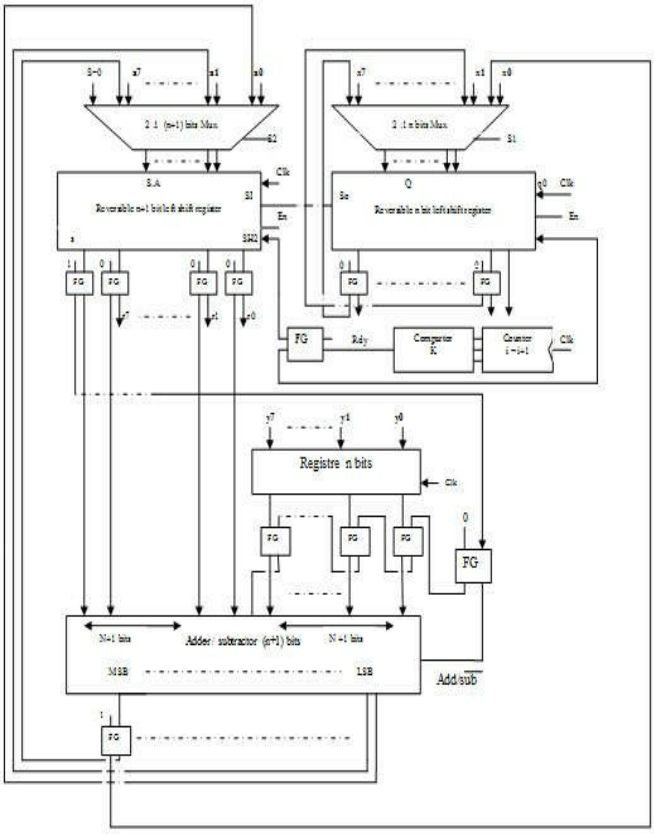


Fig 10. Proposed reversible 8 bit divider circuit

Fig 10 shows the proposed reversible divider circuit. It has 2 PIPO (parallel input-parallel output) left-shift registers. One is n+ 1 bits (n=8) named as S.A and other is n bits named as X. It equally contains an n bits register in order to store the divisor.Initially$S.A_{n-1}A_{n-2}.....A_0$=0,X($X_{n-1}X_{n-2}.....X_0$) = dividend, Y ($Y_{n-1}Y_{n-2}.....Y_0$) = divisor and rdy = 0. If the division process is completed then, the register X ($X_{n-1}X_{n-2}.....X_0$) contains the quotient and S.A ($S.A_{n-1}A_{n-2}.....A_0$) includes the remainder. On the other hand, when S1 = 0 then the two-input n-bit MUXselects dividend X ($X_{n-1}X_{n-2}.....X_0$), and if S2 = 0 then the two-input(n+1)bitMUXselects S=0andA($A_{n-1}A_{n-2}.....A_0$)= 0.Duringtheclockpulsewhen E=1,SH1=0andSH2=0, the output data from n bit MUX and (n+1)-bit MUX are loaded into X and S.A respectively. When E = 0, both S.A and X act as left-shift registers. Initially the value of S is not important, it is important just after the left shift of S.A&X (&=Concatenated), (S.A&X means SO of register X is connected to SI of S.A). If S is high ('1') then S.A-Y is performed, else, S.A+Y is computed. The complement of the MSB (most significant bit) is loaded into q0 bit position of register X. In addition, the sum is loaded into register S.A during next clock cycle (S1= 0). It includes 2n+1 clock signals to store the quotient value into register X. Finally, after 2n+1 clock signal, X stores the quotient and A stores the remainder indefinitely.

V.

D. DIVISION PROCESS EXAMPLE(14/2)

TABLE3. Division process example

Elementary operation	S.A3A2A1A0	X3X2X1X0
loading registers	00000	1110
Shift	00001	110-
Subtractor	00001 11101 1111	1100
Shift Adder	11111 000100 0001	100- 1001
Shift Subtractor ❖	00011 11101 1 00001	001- 0011
Shift Subtractor	00010 11101 1 00000 Final remainder	011- 0111 Quotient

Evaluation of the proposed reversible divider

Table 4 shows the characteristics of the proposed reversible division (n bits) without control circuit (counter, comparator).

TABLE 4. Characteristics of the proposed divider circuit.

Reversible components	Constant inputs	Quantum cost	Garbage outputs	Hardware Complexity HC (n=1)
n bit MUX 2 :1	0	4n	n	2a+ 4þ+36
(n+1) bit MUX 2 :1	0	4n+4	n+1	4a+ 8þ+66
(n)-bit register	n	5n	n+1	4a+ 6þ+46
n-bit left shift register	3n	15n	3n+2	10a+14þ+106
(n+1)-bit left shift register	3n+3	15n+15	3n+5	20a+28þ+206
(n+1)-bit parallel adder/ subtactor	n	6n+2	2n+2	7a+ 2þ
Feynman	2n+5	3n+3	0	6 a

- n-bit reversible MUX: number of garbage outputs = n, quantum cost = 4n, constant inputs=0 and the hardware complexity 2a+4þ+36.
- ❖ n+1-bitreversibleMUX:numberofgarbageoutputs = n+1, quantum cost = 4n+4, constant inputs=0 and the hardware complexity 4a+ 8þ+66.
- ❖ n-bit register: number of garbage outputs = n+1, quantum cost = 5n, constant inputs=n and the hardware complexity 4a+6þ+46.
- ❖ n-bit left shift register: number of garbage outputs = 3n+2, quantum cost = 15n, constant inputs=3n and the hardware complexity 10a+14þ+106.
- ❖ (n+1)-bit left shift register: number of garbage outputs = 3n+5, quantum cost = 15n+15, constant inputs=3n+3 and the hardware complexity 20a+ 28þ+206.
- ❖ (n+1)-bit parallel adder/ subtactor: number of garbage outputs = 2n+2, quantum cost = 6n+2, constant inputs=n and the hardware complexity 7a+ 2þ.
- ❖ Other gates (Feynman gate): number of garbage outputs = 0, quantum cost = 3n+3, constant inputs=2n+5 and the hardware complexity 7a+2þ.

VI. COMPARISON

In this section, we have compared the performance of the proposed design with some existing designs as depicted in table 4.

TABLE4.
Comparison

Conception	Constant inputs(n=8)	Quantum cost(n=8)	Garbage outputs (n=8)	Hardware complexity(n=1)
This work	88	440	99	53a + 62þ + 436
Existing design in [16]	89	454	103	65a +74 þ+526
Existing design in [10]	91	538	106	59a + 67þ + 336
Existing design in [13]	103	674	116	84a + 82þ + 426
Existing design in [14]	161	846	181	91a + 98þ + 506

Table 4 gives a comparison between the existing and the proposed 8 bits division designs. The proposed design reduces the garbage output, quantum cost, constant inputs as well as the hardwarecomplexity.

VII. CONCLUSION AND FUTUREWORK

In this manuscript, we proposed one approach for designing reversible division circuits. Then we have optimized our design. We have compared this proposed design with the existing designs. The proposed module for non-restoring division has better performance in terms of constant inputs, garbage output, and quantum cost as well as the hardware complexity than that of existing designs. In the future works we will design complete reversible ALU, develop reversible hardware description language and reversible synthesis tools.

REFERENCES

[1] Dongkun Shin, and Jihong Kim, "Intra-Task Voltage Scheduling on DVS-Enabled Hard Real-Time Systems", IEEE transactions on computer-aided design of integrated circuits and systems, VOL. 24, NO. 10, pp 1530-1549, October2005.

[2] Bishop Brock and Karthick Rajamani, "Dynamic Power Management for Embedded Systems". IEEE international SoC conference, pp 17-20, September 2003.

[3] Anantha P. chandrakasan and robert W. Brodersen, "Minimizing power consumption in digital CMOS circuits". IEEE, vol 4, pp 498-523, April 1995.

[4] Dushyant Kumar Soni1, Ashish Hiradha, "Dynamic Power reduction of synchronous digital design by using of efficient clock gating technique", International Journal of Engineering and Techniques Volume 1, pp 18- 23, May2015.

[5] Vikas, Deepak, "A review on glitch reduction techniques", International Journal of Research in Engineering and Technology. Volume: 03 pp 145-148, Feb-2014.

[6] Dhananjay E. Upasani, Sandip B. Shrote. "Standby Leakage Reduction in Nanoscale CMOS VLSI". International Journal of Computer Applications Volume 7– No.5, September2010.

[7] Onkar Nath Sabran, Shrikant Ahirwar, "An Extensive Review on Reversible Logic in VLSI", International journal of scientific progress and research. Vol-01,2014.

[8] R. Landauer, "Irreversibility and heat generation in the computing process". IBM Journal of Research and Development, vol. 5, pp. 183- 191,1961.

[9] C.H.Bernet, "Logical Reversibility of computation", IBM Journal of Research and Development, pp 525-532,1973.

[10] Noor Muhammed Nayeem, Md. Adnan Hossain, Md. Mutasimul Haque, Lafifa Jamal, Hafiz M. Hasan. Novel Reversible DivisionHardware,
IEEE, pp 1134-1138. 2009

[11] Rohini H Dr. Rajashekar S Dr. PriyatamKumar. " Design of basic sequential circuits using reversible logic". IEEE, pp 2110-2115.2016

[12] Deeptha A, Drishika Muthanna, Dhrithi M, Pratiksha M, B S Kariyappa, " Design and optimization of 8 bit ALU using reversible logic", IEEE, pp 1632-1636. 20-21May.2016.

[13] Faraz Dastan1 and Majid Haghparast," A novel nanometric fault tolerant reversible divider". International Journal of the physical Sciences Vol. 6(24), pp. 5671-5681, 16 October,2011.

[14] Faraz Dastan1 and Majid Haghparast. "A Novel Nanometric Reversible Signed Divider with Overflow Checking Capability". Research Journal of Applied Sciences, Engineering and Technology vol 4(6) pp 535-543, 2012

[15] Pallav Gupta Abhinav Agrawal, and Niraj K. Jha. " An algorithm for synthesis of reversible logic circuits" IEEE, pp 2317-2330. November 2006.

[16] A. Bolhassani and M. Haghparast, "Optimised reversible divider circuit", Int. J. Innovative Computing and Applications, pp 13-33,2016.

[17] Nagamani A, Jayashree H,H R Bhagyalakshmi, " Novel low power comparator design using reversible logic gates". IJCSE, pp 566-574, sep 2011

[18] A. Majumder, P.L. Singh, Efficient design and analysis of N-bit reversible shift registers. Procedia Computer Science, pp 199 – 208, 2015

[19] Deeptha A, Drishika Muthanna, Dhrithi M, Pratiksha M, BS Kariyappa , "Design and optimization of 8 bit ALU using reversible logic", IEEE, pp 1632-1636. May, 21, 2016

[20] Subramanian Saravanan, Ila Vennila, Sudha Mohanram, Design and Implementation of efficient reversible arithmetic and Logic Unit. SciRes, pp 630- 642.2016.

[21] Robert Wille, Anupam Chattopadhyay, Rolf Drechsler. From Reversible Logic to Quantum Circuits: Logic Design for an Emerging Technology, IEEE, pp 268-274.2016

Design of an Energy Efficient IoT enabled Smart System based on DALI network over MQTT protocol

[1]K. Premalatha

[1]PG Student, Dept. of ECE, MRCE, Hyderabad, India, Email: premalathakotthapeta124@gmail.com

ABSTRACT- The term "Internet of Things" or "IoT" refers to a hot and virgin area that is gaining importance day by day with increasing advancement in technology. With the help of micro-controllers like Arduino and micro- computers like Raspberry Pi, cheap devices can be used to measure sensor data and send it over the internet. Also, DALI (Digital Addressable Lighting Interface) is a new module of communication between the electrical equipment and a server or computer that will result in an effective realization of a smart-grid. This paper covers the communication of a DALI network of a group of lights with a Raspberry Pi (broker) over MQTT (Message Queuing Telemetry Transport) protocol keeping the concept of IoT in the background. This method proved to optimize electricity usage by optimal controlling the intensity of LEDs at various floors of the CDAC smartbuilding.

Key words – Internet of Things (IoT), DALI, MQTT, Paho C Client Library, Smart Grid, Raspberry Pi.

I. INTRODUCTION

In line with the mainstream technological innovations the emerging IoT technology aims at improving the quality of life of general masses and the efficiency of the civic amenities in a sustainable fashion while optimising economic investments by private and public sector. The IoT applications can be found in various vital fields such as: energy, health, transportation, environment, etc. Millions of application in all these fields can work optimally if there is a strong interconnection among IoT enabled devices [1]. Interconnection among IoT enabled devices via communication channels and protocols is not only a simple technological subject matter but it interests also other characteristics such as privacy, standardization, legal issues, etc. [2]. This unsurprisingly induces newer innovative challenges in IoT environment, which inspires industrial and academics researchers to go further in IoT sustainability research [3]. Confidently, the technological interfacing and communications among the IoT enabled devices (things) is essential to empower the IoT environment [4]. This is the reason why few vital functions will be performed by few key technological components and techniques, which have evolved to guarantee interconnection among heterogeneous devices adapting to the usage of very less supplies in terms of both computational time, memory, functionalities and energy resources.

This paper offers an analysis report of a small project done by the authors at CDAC, Pune to configure the electrical arrangement of a building in a smart manner using the concepts of DALI and IoT.

Accordingly, the rest of the paper is organized as follows: Section II identifies the basic apparatus required for the project and a brief description about the various protocols and modules that have been exhausted in the project. Section III provides an overview of the method through which the project was carried out and the working principles of the project. Section IV presents a brief of the results achieved from the project and an analysis of how smart grids are efficient investments for people in the time to come. Section V consists of the conclusion and outlook. Following the conclusion is a short note of acknowledgement by the authors to the people who have helped during the course of this project.

II. SYSTEM MODELLING

A. Hardware and Software used

In order to carry out this project, we made use of the following set of hardware:

- A Raspberry Pi 3 Model B
- DALI LED drivers, preferably LCM 60 DA.
- Group of LED lights

The software comprised of:

- Linux/Debian terminal

- Eclipse Paho client library for MQTT communication
- Any C programming IDE
- MQTT mosquitto broker

Before getting into the process of performing the project, it is imperative to acquaint ourselves with the basic concepts behind the modules and protocols in use:

1) MQTT: MQTT is a lightweight messaging protocol that provides resource-constrained network clients with a simple way to distribute telemetry information. The protocol, which uses a publish/subscribe communication pattern, is used for machine-to-machine (M2M) communication and plays an important role in the Internet of Things (IoT). MQTT allows devices to send (publish) information about a given topic to a server that functions as an MQTT message broker. The broker then pushes the information out to those clients that have previously subscribed to the client's topic. To a human, a topic looks like a hierarchical file path. MQTT is a good choice for wireless networks that experience varying levels of latency due to occasional bandwidth constraints or unreliable connections.

Should the connection from a subscribing client to the broker get broken, the broker will buffer messages and push them out to the subscriber when it is back online. Should the connection from the publishing client to the broker be disconnected without notice, the broker can close the connection and send subscribers a cached message with instructions from the publisher.

There are many platforms of MQTT protocol, the most popular being the HiveMQ and the Mosquitto platforms.

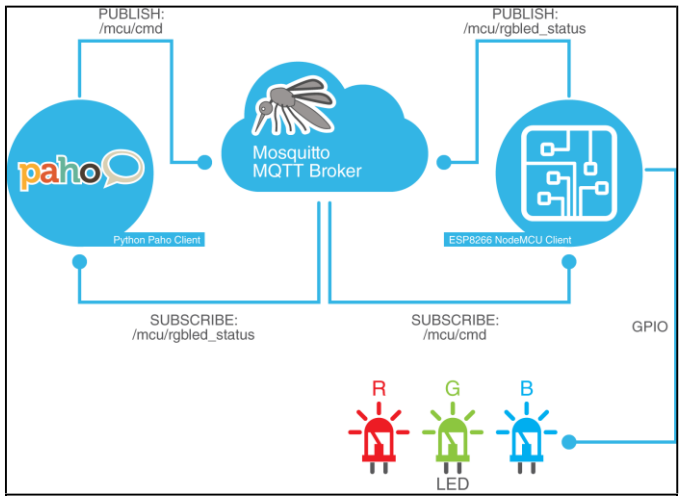


Fig. 1 Interaction of MQTT broker with the publisher and subscriber. [5]

In Fig.1, a Mosquitto MQTT broker acts as the server for communication between two clients (ESP8266 Node MCU and Paho Python), both of which can act as either a subscriber or a publisher. However, at a time, one will act as the publisher while the other will be the subscriber. This communication happens over a topic which is common to both ends. This maintains the singularity in communication over the same broker among various clients without any interference .

2) Eclipse Paho: The word ***pāho*** is a Maori word that means *to broadcast, make widely known, announce, disseminate, transmit*. The Paho project was created to provide scalable open-source implementations of open and standard messaging protocols aimed at new, exisiting, and emerging applications for Machine-to-Machine (M2M) and Internet of Things (IoT). Objectives include effective levels of decoupling between devices and applications, designed to keep markets open and encourage the rapid growth of scalable Web and Enterprise middleware and applications. Paho initially started with MQTT publish/subscribe client implementations for use on embedded platforms, and in the future will bring corresponding server support as determined by the community [6].

Paho client libraries can be written in many programming languages, viz. C, C++, Java, Python, etc. For this project, we performed Paho client programming in C.

3) DALI: DALI is a data protocol and transport mechanism that was jointly developed and specified by several manufacturers of lighting equipment. The common platform of DALI enables equipment from different manufacturers to be connected together. DALI network consists of a controller and one or more lighting devices (e.g., electrical ballasts, LED drivers and dimmers) that have DALI interfaces [10]. The controller can monitor and control each light by means of a bi-directional data exchange. The DALI protocol permits devices to be individually addressed and it also incorporates Group and Scene broadcast messages to simultaneously address multiple devices (e.g., "Group 1 goto 100%" or "Recall Scene 1"). DALI devices include LED drivers, fluorescent HF ballasts, low voltage transformers, PE cells, motion detectors, wall switches and gateways to other protocols. There can be up to 64 DALI devices on a single DALI network. Sites requiring more than 64 devices are implemented by having multiple separate DALI networks, each with up to 64 devices. These separate networks are then linked together with DALI gateways and a data bus running a high level protocol.

III. METHODOLOGY

This project works on the basic MQTT protocol, i.e. a publisher sends some data to an MQTT broker (created on the Raspberry Pi itself) which is further transferred to the subscriber. Both the publisher and subscriber are subscribed to the same topic. Each Raspberry Pi system is capable of handling at most 64 lights (clients) as per the DALI protocol.

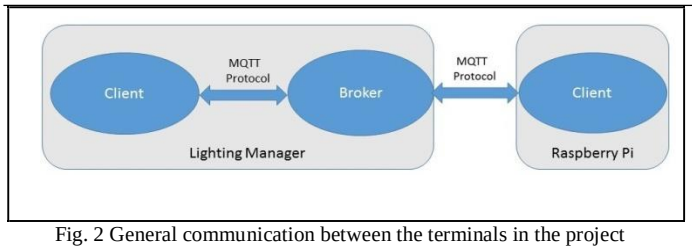
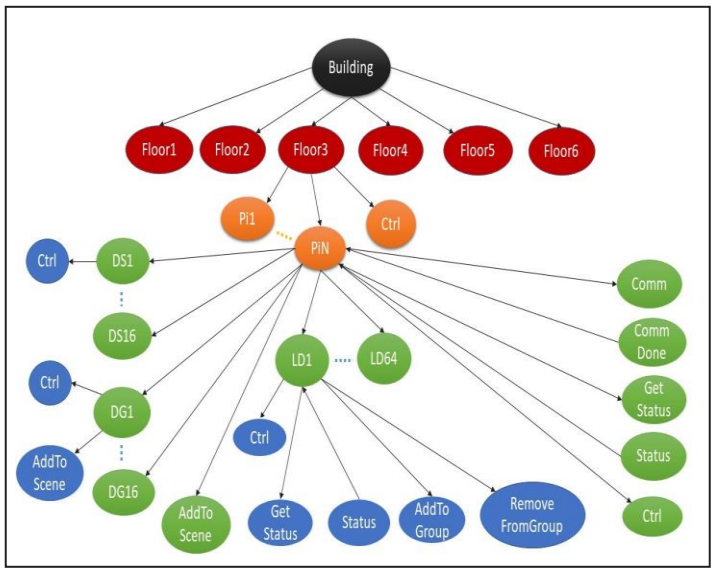


Fig. 2 General communication between the terminals in the project

Fig.2 gives a pictorial representation of the skeleton of the project where Lighting manager represents the console of DALI drivers, each of which connects to maximum LEDs



(nodes).

Fig. 3 Sequential range of topics undertaken by current research methodology for optimal energy saving with more than 64 devices DALI Network

The first stage in the project is to begin with the installation of Paho C client library on the Raspberry Pi. Thereafter, the mosquitto MQTT platform is installed. Once the installation process is over, the structure of topics is brought into existence. As is evident from Fig.3, the range of topics available is diverse. Each topic corresponds to a different set of topic. The hierarchy of topics is as follows:

a) *Building -> Floor n -> Pi n*: Each floor can have many Raspberry Pis to drive numerous DALI drivers. Hence, once the communication happens over this topic, the user has the option to control the state of either one or more than one LEDs lights on that floor through the corresponding Pi.

b) *Building -> Floor n -> Ctrl*: If the user wishes to control the entire floor and all Raspberry Pis and DALI drivers on that floor, this topic will be sent.

c) *Building -> Floor n -> Pi n -> Comm*: This topic is sent by the publisher to start the communication with the subscriber (DALI drivers).

d) *Building -> Floor n -> Pi n -> Comm Done*: This topic returns a callback signal to the broker whether successful communication has been formed with the client. If not, the broker will retry communication with the client until it's done.

e) *Building -> Floor n -> Pi n -> Get Status*: If the user wishes to know the current brightness status of the lights or of a particular DALI driver, this topic will be communicated with.

f) *Building -> Floor n -> Pi n -> Status*: The client returns its current status to the user as a return call to the above topic.

g) *Building -> Floor n -> Pi n -> Ctrl*: The user can get control over a single Pi on a particular floor and control all the DALI drivers with a single command through this topic.

h) *Building -> Floor n -> Pi n -> DS N -> Ctrl*: The term DS in this topic implies 'DALI Scene' and N represents the particular DALI driver that has 4 nodes. This topic enables the user to only control the brightness scene and make changes to it, perhaps like moderating the dim and bright lighting according to the situation. For instance, if it is a projection room, the lights are bright initially but may be automatically dimmed by the user once the projection of a presentation or movie begins in the room.

i) *Building -> Floor n -> Pi n -> DG N ->Ctrl*: The term DG here signifies 'DALI Group' and N is same as in the above topic. Here, there are total 16 DGs under each Pi and each DG consists of 4 nodes. We can control the settings and state of each group here, however cannot change the scene settings.

j) *Building -> Floor n -> Pi n -> DG N -> Add to Scene*: This topic is useful when the user wishes to access a particular DG and add some brightness settings to its current scene without altering it.

k) *Building -> Floor n -> Pi n -> Add to Scene*: Instead of changing the scenes of each DG or DS, one can also add settings to the scene of all 64 nodes under one Pi through this topic.

l) *Building -> Floor n -> Pi n -> LD N -> Ctrl*: The term LD in this topic refers to 'LED DALI', i.e. instead of having the nodes (LEDs) as groups, they can also be individually placed under a Raspberry Pi

thereby giving access to a Pi to 64 LDs. Hence, with this topic the state of the LDs can be controlled individually.

m) *Building -> Floor n -> Pi n -> LD N -> Get Status*: Same as topic e), only difference being that an individual LD's status is being sought for this time by the user.

n) *Building -> Floor n -> Pi n -> LD N -> Status*: Similar to topic f), this one returns the status of the LD which was sought by the user in the earlier topic.

o) *Building -> Floor n -> Pi n -> LD N -> Add to Group*: There is also an option to add the individual LDs to an existing DG or into a new one, keeping in mind that maximum number of nodes in each group should not exceed 4.

p) *Building -> Floor n -> Pi n -> LD N -> Remove from Group*: The LDs can be removed from the groups that they are a part of too.

How does it work?

The above range of topics describe the vast scope of available options to the user for controlling and monitoring a network of lights either in a home or even in a building. However, this is achieved through programming in Paho C client library.

We program both ends of the broker in such a way that firstly a client is created. A user-friendly environment is developed where the topics are listed on the screen, and the user can select a topic of his choice. Once the topic is also initialized, the subscriber client however tries connection with all the topics. Depending on the topic that the publisher has connected to, the subscriber reaches out to the set of topics corresponding to that topic and returns its status to the publisher via the broker.

Depending on the choice entered by the user, the corresponding topic gets generated on the publishing side. This choice of topic is sent to the MQTT broker (Raspberry Pi), which further channelizes the communication of data with the subscriber over this topic. The subscriber however generates all possible topics and tries to connect to each one of them at very short intervals after creating the client. The communication takes place when the topics of both publisher and subscriber match with each other. At this instant, the message from the publisher is published to the subscriber, and the resulting set of operations corresponding to that topic are performed. Eventually, we finally see the result physically with the varying brightness level of the lights or LEDs. The same process continues and various publishers can publish to the same subscriber over different topics. Also, the connection process is made in such a way that even if the connection is lost or broken from either sides, the client tries to reconnect with the broker and does not disconnect, unless disconnected manually by the user.

IV. RESULTS

Although we approached this project at its beginning stage, yet by implementing the proposed methodology many results were realized by us towards the end of it. The terminal windows consisting of the publishing and subscribing operations has been shown in section III of this paper.

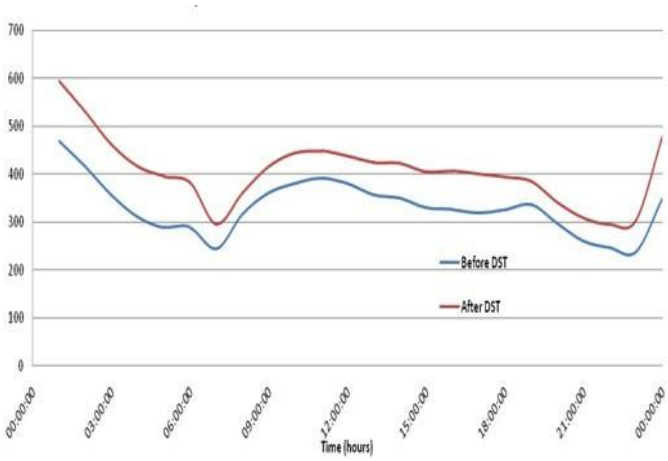


Fig. 4 Graphs of energy saving (in kWh) on Y-axis before and after DALI Scheme Transfer (DST) in a day

The analytics of this shows that on the peak periods of usage, the load can be increased and decreased vice versa via DALI controller. Also, the user desired topic can be sent to the MQTT broker (Raspberry Pi), which further publishes the data to the subscriber for optimally setting the brightness of LEDs as shown in Fig. 5a and 5b. This causes the energy saving in the electricity usage as shown in Fig. 4, and this will play a huge role in the energy saving in urban areas paving the way for a smarter demand response on consumer side.

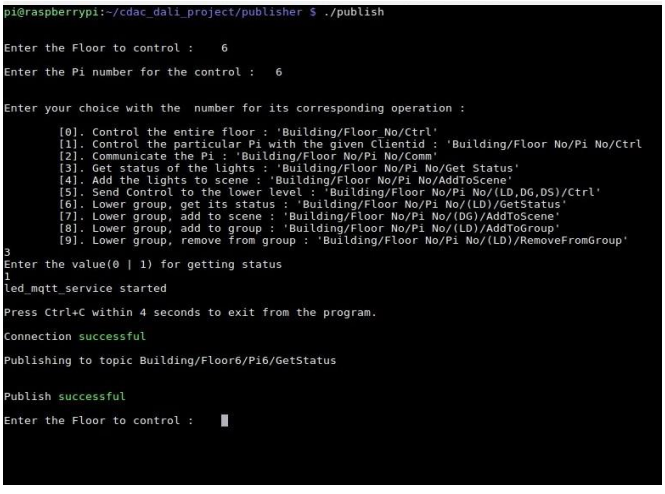


Fig. 5a Terminal window for the publisher

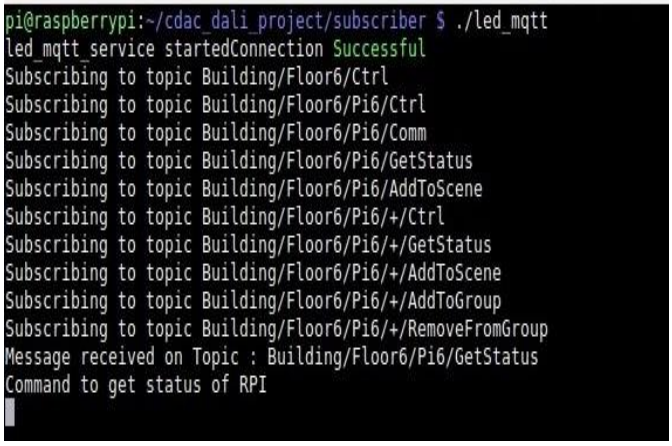


Fig. 5b. Terminal window for the subscriber

V. CONCLUSION

The coming decade is going to witness a revolutionary change in the manner of usage of energy and smart monitoring systems are going to come in vogue. We have been able to successfully implement a smart energy saving methodology via communication of a DALI network with a Raspberry Pi over MQTT (Message Queuing Telemetry Transport) protocol keeping the concept of IoT as the background for optimizing the brightness control of LEDs at different floors in the smart building of CDAC Pune. The constraint of using only 64 DALI devices on a single DALI network was overcome by having multiple separate DALI networks, each with up to 64 devices in the entire set up with reduced delay and maximum energy saving.

REFERENCES

[1] aheeruddin and Munish Manas, “A New Approach for the Design and Development of Renewable Energy Management System through Microgrid Central Controller”, Energy Reports, Elsevier Inc., vol. 1, pp. 156-163, 2015.

[2] https://en.wikipedia.org/wiki/Digital_Addressable_Lighting_Interface

[3] Zaheeruddin; Manas, M., "Development of Self-Sustainable Technologies for Smart Grid in India," in Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on , vol., no., pp.563-568, 13-14 Feb. 2015.

[4] Zaheeruddin and Munish Manas “Analysis of Design of technologies, tariff Structures and regulatory policies for sustainable growth of the Smart grid” in Taylor and Francis’s “Energy Technology and Policy” Journal, Volume 2, issue 1, pp. 28-38, 2015.

[5] B. Saharia and Munish Manas, “Viability analysis of Photovoltaic/Wind hybrid standalone Distributed Generation based system as applicable in isolated community of north-east region of India”, Distributed Generation and Alternative Energy Journal , vol. 32, issue 1, PP. 49-80 • January 2017.

[6] W. Wu, M. Wu and Y. Liu, "A Design of Embedded DALI Controller," *2006 4th IEEE International Conference on Industrial Informatics*, Singapore, 2006, pp. 1237-1240.

[7] doi: 10.1109/INDIN.2006.275816

Verification and Simulation of New Designed NAND Flash Memory Controller

Akarapu Natraj Renu

PG Student, Dept. of ECE, MRCE, Hyderabad, India, Email: natrajrenu@gmail.com

Abstract— **In this paper a NAND flash memory controller was designed. For the best use of NAND type flash memory we design a new Arithmetical and Logical Unit (ALU) for calculating increment, addition, subtraction, decrement operations etc. In this memory controller we design single memory cell, memory module, a decoder etc. These all are encapsulated inside a controller and this is on top most in hierarchy. NAND flash memory is a non volatile storage media used in today daily life electronic equipments. NAND flash memory is programmed on page by page basis. Typically programming time is very less few micro second per page. This NAND flash memory controller architecture can be used with a real secure digital card, multimedia card (SD/MMC), digital cameras etc. The NAND Flash memory controller can be an internal device, built into the application processor or host, or designs can incorporate an external, stand-alone chip. Experimental results show that the designed controller give good performance and full fill all the system specifications. We have used FPGA chip for download our code.**

Keywords-Flash memory; Non Volatile; Arithmetical and Logical Unit (ALU); Encapsulation; Field Programmable Gate Array (FPGA);Secure Digital card/Multi Media Card (SD/MMC);Hard Disk Drive (HDD);

I. INTRODUCTION

Flash memory is being widely used as a storage medium in mobile devices because of its low power consumption, small form factor, and high resistance to shock and vibration. As the density of a flash memory chip increases and the price continues to drop, the flash memory is being adopted in more diverse storage applications. Flash memory is the combination of two technologies-EPROM and EEPROM. The term "Flash" means - A large chunk of memory (memory cell) could be erased at one time". On the other hand in EEPROM each byte is erased one by one manner.

Flash memory has characteristics that are different from conventional storage devices such as HDD. Thus, specialized hardware and software are required to use flash memory as a storage device. The role of flash memory software is particularly important because it has to deal with the peculiarities of flash memory. It also needs to be diverse because flash memory is used in a wide spectrum of applications ranging from micro-embedded systems (e.g., a sensor-node) to large-scale servers [5]. Flash memory controller offers higher capacity for fast data transfer and random access of memory in I/O operations [1]. It is possible to design a simple memory mapped interface to hardware with NAND flash memory. NAND flash memory controller has also a bidirectional bus in between peripheral devices and controller, controller internal bus for various functional blocks and controller with memory connections. For the improvement of product lifetime and system performance we always design an excellent NAND flash memory controller. Single Level Cell (SLC) and Multi Level Cell (MLC) two techniques are used for storing data in memory cells. SLC offers 100,000 erase program cycle while MLC offers about 10,000 erase program cycle [2].

NAND flash memory controller has also a bidirectional bus (Data, Address, and Control). NAND flash cell are placed together for saving 60% cell size over NOR flash cells. NAND flash memory controller provides a serial access of data blocks in a very high speed [6]. Single Level Cell (SLC) and Multi Level Cell (MLC) two techniques have been used for storing data in memory cells. A software called Flash Transaction layer (FTL) use for wear leveling and bad block management technique. All vendors provide FTL software [9].

The pipelining and parallel processing concepts are applied for systematic design approach of a systolic array processor. The systolic array architecture and iterative Very Large Scale Integration (VLSI) architecture is applied for good performance. It makes the circuit design easy for implementation [8].

II. NAND FLASH MEMROY CONTROLLER ARCHITECHTURE

The capacity of NAND flash device is improved day by day, architecture are also improved day by day. Latest overall structure of NAND flash device is looks very similar to its conventional structure. A NAND flash controller implements memory mapped interface [10]. A multiplane array packs contains its own set of Cache/Data registers, more memory cells on a die and partitioned it into several plans. In multiplane array packs all operations performs parallels. Inside a single flash array operation time multiple pages of data can be programmed, read, write, fetched etc. so average data access time is small. In multiplane commands some addresses are restricted.

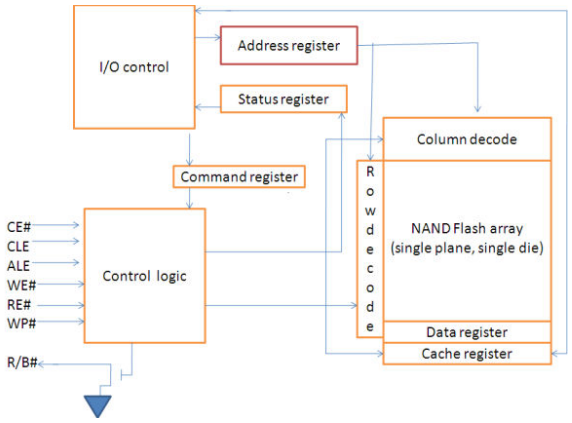


Figure 1. Block Diagram of a NAND Flash Memory Controller

NAND flash controller has flash chips. In today life it is widely used in storage devices. One more recent application for flash memory is as a replacement for hard disks. Flash memory does not have the mechanical limitations and latencies of hard drives, so a Solid State Drive (SSD) is attractive when considering speed, noise, power consumption, and reliability. It always follo ws Open NAND Flash Interface (ONFI) standard [3].

In NAND flash device we have mainly I/O control block, control logic, NAND flash array. A NAND flash array includes two dimensional NAN D flash cells, Row/Column address decoder and cache / data registers [11]. It has a shared, multiplexed, bidirectional (command, address and data) I/O bus. Figure-1 shows the block diagram of a NAND Flash Memory Controller. A multiplane array packs contains its own s et of Cache/Data registers, more memory cells on a die partitioned into several plane. In multiplane array packs, a ll operations are performed in parallel. Thus, inside a si ngle flash array, multiple pages of data can be programm ed, read, write, fetched, so average data access time is reduced. In multiplane commands some addresses are restricted. So new NAND flash chip with multidie and multiplane support is always increase performance, reduce the data access average time, and increase parallel execution of commands

[7]. NAND flash devices are programmed on a page by page basis. Typically programming time is a fe w hundred micro second per page. NAND flash cell can be programmed and erased only for limited time period (100,00 0 times for SLC and 10,000 times for MLC) before it fails. To improve this limitation, flash memory performance has b een increased by using wear leveling technique. The technique spreads the memory cell use evenly to different physi cal pages. So the entire flash device is used equally to im prove the life of flash memory [2]. Bad block manageme nt technique and wear leveling technique use some remapping technique of logical to physical address of the memory device. They all provide an FPGA to facilitate the implementation of a wide range of NAND Flash Memory Controller [4].

III. SIMULATION AND RESULT

A. Single Micro cell Module

In the simulation result, the NAND flash memory cell is simulated using Xilinx ISE Software and modelsim simulator. As shows in Figure 2, the RTL view of NAND Flash memory cell has been gene rated after the synthesis. The modelsim waveform is displayed on Figure 3 indicating the write and read operation. Write operation takes place only when both word line and write enable are '1'. The q and qbar line are modified when this situation occurs as seen in the waveform and Read operation takes place when both word line and read enable sh uld be high so that read out shows the data stored in q.

Figure 2. RTL Schematic View of Single Micro Cell

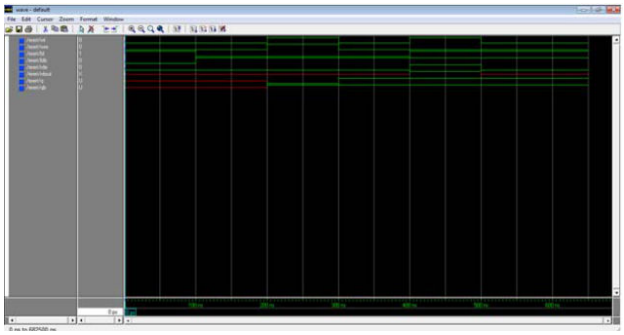


Figure 3. Simulation Waveform of Single Micro Cell

B. Full Adder

The simulation of full adder carried using Xilinx ISE Software and modelsim is shown in Figure 4. The RTL view of full adder was generated after the synthesis and it displays the internal architecture of full adder. The modelsim waveforms are displaye d in Figure 5 and depict the sum and carry output. Sum tak es place only when any one input is high and carry operation is done when at least two inputs are high.

Figure 4. RTL Schematic View of Full Adder

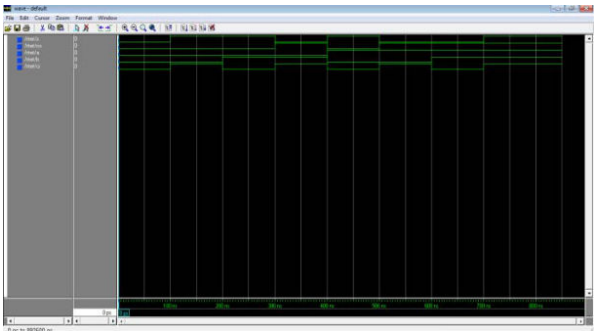


Figure 5. Simulation Waveform of Ful l Adder

C. Arithmetic and Logical Unit (ALU)

In the simulation, results of Arithmetic and Logical Unit (ALU) have been used in Xilinx ISE Software and modelsim simulator as shown in Figure 6. The RTL view of ALU has been generated after the synthesis and it displays the internal architecture of ALU. The mod elsim waveform is displayed in Figure 7 which indicates all mathematical operations. The addition operation takes place only when op_sel. Operation select lines are zero a nd provide input value to the dat_a and dat_b and chec k the results on dat_out as well as all flag register like carry flag, Sign flag, auxiliary carry flag, parity flag, etc. Thus all mathematical operation can be carried by changing the value of op_sel.

Figure 6. RTL Schematic of AL U

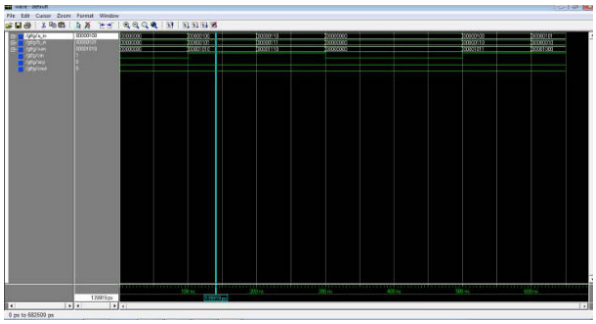


Figure 7. Simulation Waveform of ALU

D. Memory Module

In the simulation, result of Me mory Unit using Xilinx ISE Software and modelsim simulator are shown in Figure 8. The RTL view of memory unit is generated after the synthesis to display the internal architecture of Memory Unit. The modelsim waveform is displayed in Figure 9. It indicates how the data can store in to the memory with the help of read and write operation.

Figure 8. RTL Schematic of Memory Module

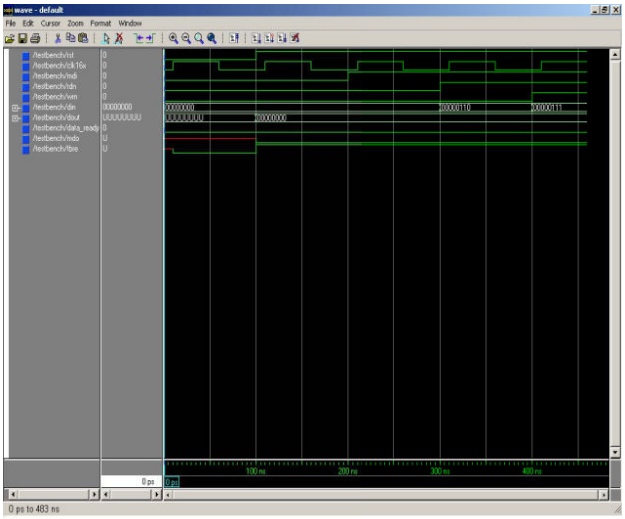


Figure 9. Simulation Waveform of Memory Module

E. Memory Read Cycle

In Figure 10, the RTL view of Memory Read Cycle has been generated after the synthesis to display the internal architecture of Memory Read Cycle. The modelsim waveform is displayed in Figure 1 1. During Memory Read Cycle when reset = '1' then both the mode operation are zero and when apply the clock pulse the data at ¼ and ¾ points in data cell is generated and when increment the clock the data are in a position to g o from serial to parallel conversion. After some time it will also shift to data register. In this way, data are shifted to the particular data register and finally read the data at that location.

Figure 10. RTL View of Memory Read Cycle

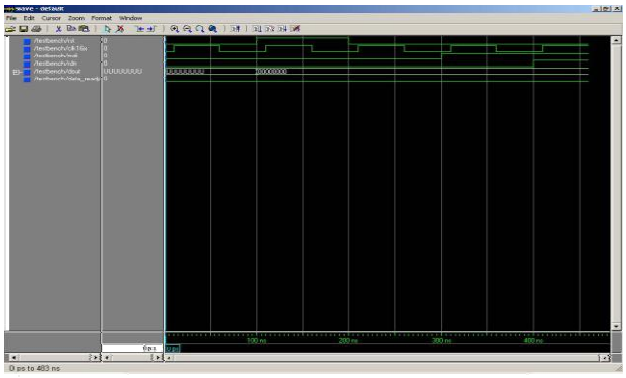


Figure 11. Simulation Waveform of Memory Read Cycle

F. Memory Write Cycle

In Figure 12, the RTL view of Memory Write Cycle has been generated after the synthesis to display the internal architecture of Memory Write Cycle. The modelsim waveform is displayed in Figure 13. During Memory Write Cycle when reset = '1' then both the wrn1 and wrn2 operation is one and when enable the clock detect the edge on write pulse. After some time, detect edge on write pulse to load transmit buffer. When increment the clock the data are in a position to go from transmit shift register to transmit buffer. In this way, data are shifted to the particular data register and finally write the data at that location.

Figure 12. RTL Schematic of Memory Write Cycle

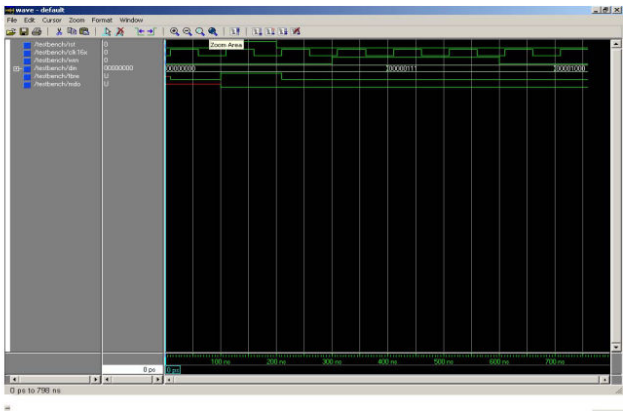


Figure 13. Simulation Waveform of Memory Write Cycle

G. NAND Flash Memory Controller

Figure 14 shows Xilinx RTL Schematic view of NAND Flash Memory Controller. Figure 15 are shows the Simulation Waveform of NAND Flash Memory Controller. We bind all the above components inside this module. This module is stand at the top of hierarchy. Various pins are described below-:

- nand_ale : This pin indicates NAND flash advanced latch enable when this pin is high, the NAND flash memory controller work and when it is low, the NAND flash memory controller are in latch mode.
- nand_ce : This pin indicates NAND flash chip enable. When this pin is high the NAND flash memory controller are in working mode, it reflects that the data is storing in a memory in sequential order. When it is low, the NAND Flash memory controller do not worked.
- nand_cle: This pin indicates NAND flash clear mode. When this pin is high the data are cleared in the main memory of NAND flash memory controller. When it is low it will not worked.
- nand_re: This pin indicates NAND flash reset. When this pin is high the NAND flash memory controller are in reset mode i.e. all operations of NAND flash memory controller stopped. When it is low, it will work properly.
- nand_tri_en: This pin indicates NAND flash tri-state enable. When it is high, it will indicate the data are in tri-state i.e. high impedance state. When it is low it may work properly.
- nand_we : This pin indicates NAND flash in write operation. When it is high, data are in writing process in a write mode i.e. we can write the data at particular memory address. When it is low, it will not provide a permission to write the data.
- nand_opr<2:0> : This pin indicates selection between reset state (000), read ID state (001),read

page state (010) and write page state (011). After that state 100,101,110,111 (4 states) reserve for future extension.

- nand_rdy : This is the NAND flash ready input signal from the memory device, only accessible when high. If it is low it will not be accessible.
- prog_2,prog_3,…prog_24 :These 23 pins used for write states
- read_2,read_3,…read_21 : These 20 pins used for read states.
- rid_2, rid_3, …. , rid_9 : These 8 pins used for select ID states.
- rst_low:This pin indicates the reset input of the system. It is normally low.
- rst_1, rst_2, rst_3 : These 3 pins used for reset states.

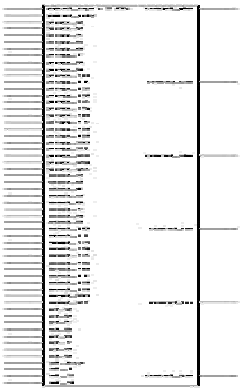


Figure 14. RTL Schematic of NAND Flash Memory Controller

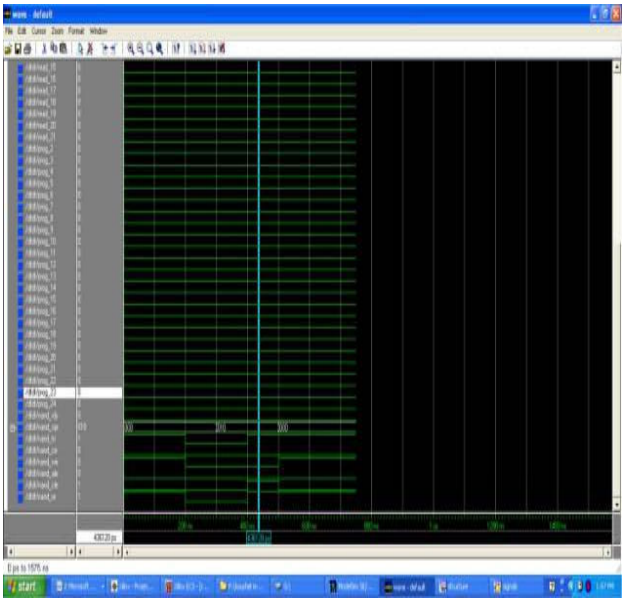


Figure 15. Simulation Waveform of NAND Flash Memory Controller

CONCLUSION

In this paper NAND flash memory controller for SD/MMC memory card using FPGA was designed. The test results show that NAND flash memory controller architecture will achieve a high performance. In the proposed NAND flash memory controller, all the blocks like microcell, microcontroller, ALU, full adder, memory unit, memory read cycle, memory write cycle have been developed. Proposed architecture can work with all the embedded computing system as a replacement of conventional Hard Disk Drive (HDD) with a very huge size of NAND flash memory. In this controller we design only four states reset state, read ID state, read page state, write page state. In future extension we have four reserve states.

ACKNOWLEDGMENT

This work was supported by ShriRam Collage of Engineering and Management, Banmore.

REFERENCES

[1] Shawn Kung, “Native PCIs SSD Controllers Next Generation Enterprise Architecture for Scalable I/O Performance”, White Paper Marvell Jan-2012.

[2] Yang Ou, Nong Xiao, “A Scalable Multi-channel Parallel NAND Flash Memory Controller Architecture”, 6th Annual IEEE Chinagrid Conference, Page 48-53, 2011.

[3] Yu Cai, Haratsch, “FPGA Based Solid State Drive Prototyping Platform”, 19th IEEE Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), Page 101-104, 2011.

[4] Yu Chi, Erich F. Haratsch, Mark McCatney and Ken Mai “FPGA-Based Solid State Drive Prototyping Platform,” In IEEE International Symposium on Field Programmable Custom Computing Machines 2011.

[5] Yang Ou, Nong Xiao and Mingche Lai “A Scalable Multi-channel Parallel NAND Flash Memory Controller Architecture,” Sixth Annual ChinaGrid Conference 2011.

[6] International Technology Roadmap for Semiconductor, 2007.

[7] Micron Technology, Boise, Idaho. Technical Note TN-29-17 “Design and Use Considerations for NAND Flash Memory,” 2006.

[8] Altera Corporation, Stratix data isheet, May 2003.

[9] Intel Corp., Understanding the Flash Transfer Layer (FTL) Specification, 1998.

[10] Hoeseung Jung, Sanghyuk Jung, “Architecture Exploration of Flash Memory Storage Controller through a Cycle Accurate Profiling”, IEEE Transactions on Customer Electronics, Volume: 57, Page 1756–1764, 2011.

[11] Teoh, J.N., Wong W.E., “FPGA Implementation of Nonlinear Control on Hard Disk Drive”, IEEE International Conference on Control and Automation, Page 609- 614, 2009.

ADVANCED VIDEO CODING/H.264 ENCODING AND DECODING PROCESS WITH IPPPP SEQUENCES

1Anitha Perla 2Chella Moses
1*Assistant professor, Dept. of ECE, MRCE, Maisammaguda, Hyderabad
Assistant professor, Dept. of ECE, MRCE, Maisammaguda, Hyderabad
Anithakrishna77p@gmail.com,moses.chellaece@gmail.com
India

***Abstract*—** As the world expanded around us and increased the popularity of the Internet by sending receiving, uploading or downloading the high definition videos, it was necessary to use a good technology to reduce the size of the dedicated video and specialized high-quality one. If the videos are send or receive, they need a wide bandwidth to capture this amount of information in the video. Based on the above, the H.264/AVC is a good video coding standard that gives great results for encoding and decoding videos in terms frames. This technology was developed jointly by (ITU-T) International Telecommunication Union–Telecommunication Standardization, and (ISO) International Organization for Standardization. Our work involves applying the encoding and decoding process of the standard using MATLAB TOOL. The work is focusing in inter frame prediction using the (IPPPPP) frame pattern. The video that was subjected to encoding and decoding processing is using with different video coding standards such as H.264/AVC.

Keywords—IPPPP, High Quality, video, encoding, decoding

I. INTRODUCTION

H.264/AVC is the international video coding standard. It is coding. approved by ITU-T as Recommendation H.264 [1] and by ISO/IEC as International Standard MPEG (Motion Picture Expert Group)-4 part 10 AVC[2]. It is widely used for the transmission of Standard Definition (SD) and High Definition (HD) TV signals over Satellite, cable, and terrestrial emission and the storage of high-quality SD video signals onto DVDs[3]. However, an increasing number of services and growing popularity of high definition TV are creating [4] greater needs for higher coding efficiency. Moreover, other transmission media such as Cable Modem[5] offer much lower data rates than broadcast channels, and enhanced coding efficiency can enable the transmission [6] of more video channels or higher quality video representations within existing digital transmission capacities[7]. Digital Video is a sequence of still images or frames and represents scenes in motion[8]. A video signal is a sequence of Two Dimensional (2D) images projected from a dynamic three dimensional (3D) scene onto the image plane of a video camera[9]. Video coding is the process of compressing and decompressing a digital video signal[10]. Coding of video is performed picture by picture. Each picture to be coded is first partitioned into a number of slices[11]. Slices are individual coding units in this standard as compared to earlier standards as each slice is coded independently[12]. The hierarchy of video data organization is as follows: picture-slices-macroblocks-sub macro blocks-blocks-pixels[13]. The main elements of the H.264, source coder are prediction, block transformation (spatial to frequency domain translation), quantization, and entropy[14-16].

II. LITERATURE SURVEY

The purpose of this section is to provide a short description of the various video compression standards. Standards allow interoperability between different manufacturers and a variety of equipment worldwide. Two standards bodies have strongly influenced the development of video compression standards: the International Telecommunications Union (ITU) and the Moving Pictures Experts Group (MPEG)[3-5]. The oldest body is the ITU .Of particular interest is the Telecommunication Standardization Sector, or ITU-T sector[11]. Their goals have been transmitting video over both the analog and digital telephone system although in recent years the difference between their goals and ISO MPEG are very much blurred[16]. Video compression standards developed by ITU are designated by the label"H.26*X*"[17].

In the early 1990's the International Organization for Standardization (ISO) began looking at video compression for computer and multimedia applications. ISO formed MPEG to develop video compression standards for ISO. MPEG can be thought of as the "computer guys." Standards developed by MPEG are designated by the label "MPEG-*x*." The MPEG committee tends to be better known than the ISO although both had an equally important impact on the video compression industry[18]. It is important to note that both ITU and MPEG video compression standards only describe the decoder and not the encoder structure[19]. The standards describe the syntax of the encoded bit stream as well as behavior of a compliant decoder. Developers then are allowed to design the encoder anyway they want as long as it produces a compliant bit stream[21-21].

In order to achieve better compression, the video may be pre-processed before being processed by the encoder. Since errors may be introduced to the compressed video data, error conceal mentor block-artifact reduction techniques may be used after the decoding process to enhance the overall quality of the video. It should be emphasized that such techniques are not part of the video compression standards.

A timeline of the development of video compression standards is shown in Figure 2.2, with the

contributions of MPEG and ISO shown above and below respectively. It should be noted that MPEG and ITU jointly developed two standards, these are MPEG-2/H.262 and MPEG-4: Part10/H.264.

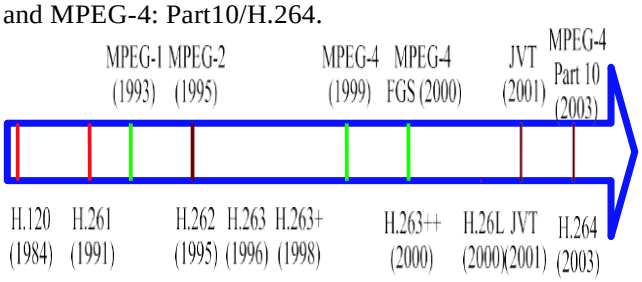


Fig. 2.2. Timeline of video coding standards. MPEG-2/H.262 and MPEG-4: Part 10/H.264 were joint projects of MPEG and ISO.

H.120 Standard:

H.120 was the first digital video compression standard. It was developed by COST 211. The video turned out not to be of adequate quality, there were few implementations, and there are no existing codecs for the format, but it provided important knowledge leading directly to its practical successors, such as H.261.The drawbacks are H.120 video was not of good enough quality for practical use it had very good spatial resolution (as differential PCM works on a pixel-by-pixel basis), but very poor temporal quality. It became clear to researchers that to improve the video quality without exceeding the target bitrate for the stream, it would be necessary to encode using an average of less than one bit for each pixel.

H.261 Standard:

H.261 is an ITU-T video compression standard, first ratified in November 1988. It is the first member of the H.26x family of video coding standards in the domain of the ITU-T Video Coding Experts Group (VCEG), and was the first video coding standard that was useful in practical term. H.261 was originally designed for transmission over ISDN lines on which data rates are multiples of 64 kbit/s. There are however a few difficult problems in H.261: Motion vector search, Bit-rate Control, Propagation of Errors.

H.262 Standard:

H.262or MPEG-2 Part 2 (formally known as ITU-T T Recommendation H.262 and ISO/IEC, also known as MPEG-2 Video) is a video coding format developed and maintained jointly by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG). MPEG-2 Video is similar to MPEG-1, but also provides support for interlaced video (an encoding technique used in analog NTSC, PAL and SECAM television systems). MPEG-2 video is not optimized for low bitrates (less than 1 Mbit/s), but outperforms MPEG-1 at 3 Mbit/s and above. All standards-conforming MPEG-2 Video decoders are fully capable of playing back MPEG-1 Video streams.

H.263 Standard:

H.263 is a video compression algorithm and protocol which is standardized by ITU. It is due to be published sometime in 1995/1996. It was designed for low bitrate communication, early drafts specified data rates less than 64 Kbits/s, and however this limitation has now been removed. It is expected that the standard will be used for a wide range of bitrates, not just low bitrate applications, and expected that H.263 will replace H.261 in many applications. The Video source coding algorithm of H.263 is based on Recommendation H.261 and is a hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy, however with some changes to improve performance and error recovery. So, In comparison with video compression H.261 which is widely used for ISDN video conferencing, H.263 can achieve the same quality as H.261 with 30-50% of the bit usage. Most of this is due to the half pixel prediction and negotiable options in H.263. H.263, in addition, is also better than MPEG-1/MPEG-2 for low resolutions and low bitrates.

H.264 / MPEG-4 Part 10: Advanced Video Coding

In early 2000 ITU and MPEG began working jointly again on a new video com- pression standard. They formed a group known as the Joint Video Team (JVT) to examine new issues in video compression . The offcial ITU and MPEGdesigna- tions are H.264 and MPEG-4 Part 10: Advanced Video Coding respectively. In this thesis, the ITU designation H.264 will be used to refer to the standard.

This standard was designed to target a wide variety of applications, including wireless , IP networks , and digital cinema.The standard has defined two separate main coding layers: the Video Coding Layer (VCL) and the Network Abstraction Layer (NAL). The VCL will be the focus of this paper, although the NAL will be discussed in some detail.

In comparison to previous standards, this standard is a departure from earlier standards, introducing many new technical features to further increase compression effciency, such as flexible macroblock sizing, 1/4-pixel interpolation , multiple reference picture capabilities, and an in-loop filter.

III.PROPOSED WORK

Encoder Decoder:

The objective of image processing is to analyze image data to make the system to understand, recognize and interpret the processed information available from the image pattern. For example, noise removal, sharpening, or enhancing brighten an image, making it easier to identify key features. These enhancement can be classified into two main groups – spatial domain based and transformation domain based methods . The popular histogram equalization (HE) based methods come in to the spatial domain, which can be further classified into two categories, i.e., global and local histogram equalization based techniques. Global Histogram Equalization (GHE) methods explore the histogram information of the entire image to form its transformation function. The classical HE can efficiently utilize display intensities, but it tends to over enhance the contrast if there are high peaks in the Histogram, which

often results in a harsh and noisy appearance of the output image.

The video can be divided in to number of frames, that sequence can me in the form of f1,f2,f3…….fn-1,fn.Then again frame can be divided in to macro block is encoded in an inter and intra mode can be perform with motion estimation and motion compensation that is to be count as prediction process.

In this work inter prediction mode is performing with motion compensation prediction with previous frame that is to be consider as reference frame. That the sequence to be consider as IPPPPPPP----.

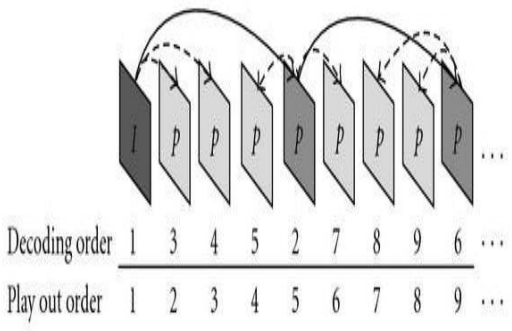


Figure. 2. IPPPP--- sequence

The prediction picture is subtracted from the current block to produce as residual block that process can be consider as transformation process. Here transformation technique has implemented with hybrid wavelet transform. And quantized to be a set of coefficients i.e quantized transform coefficients, which are to be ordered and apply entropy process.

H.264 decoder is inverse process of H.264 encoder and it is used to decode the encoded video come from the encoder. Decoder receives compressed bit stream and decodes data elements to produce set of quantized coefficients, these coefficients has to be decoded from the bit stream and decoder creates prediction block, from that prediction block it has to send to Deblocking filter and generate frames.

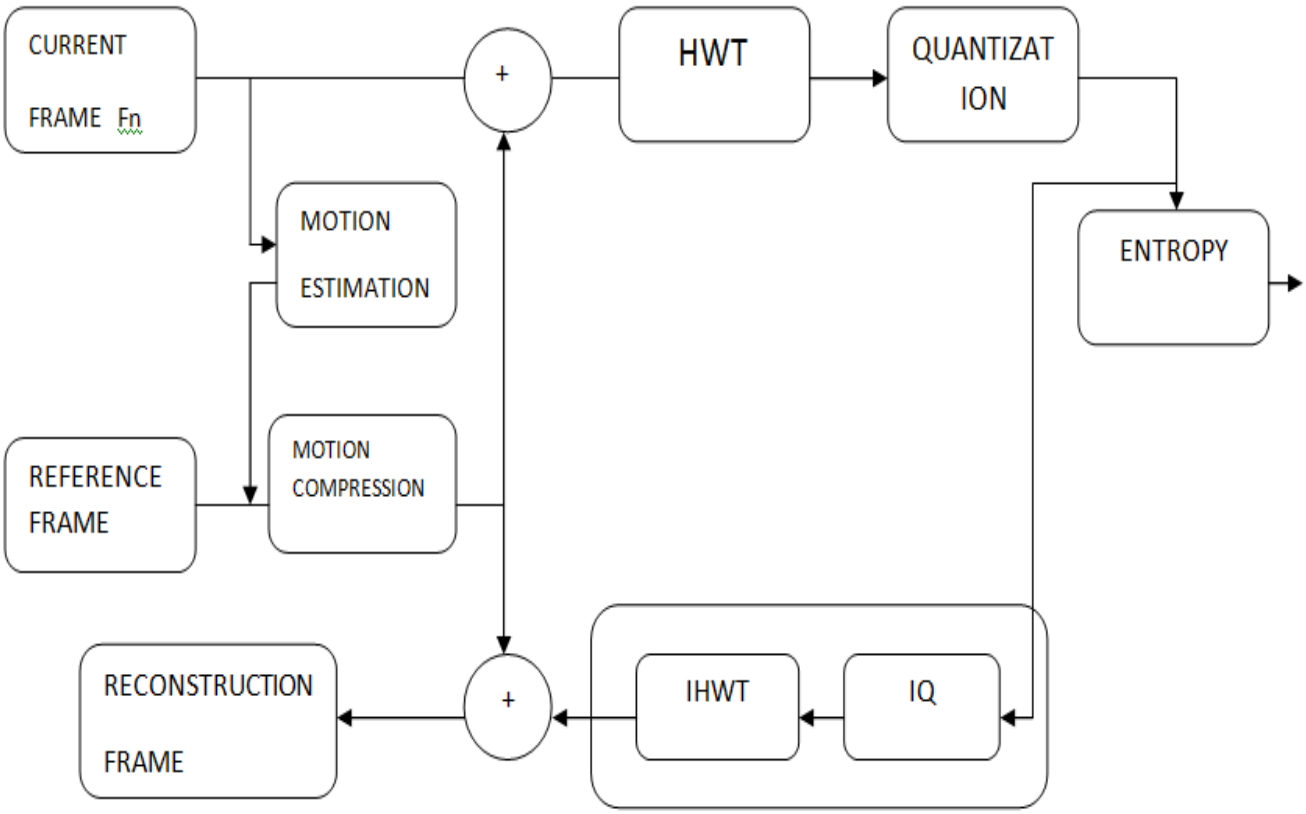
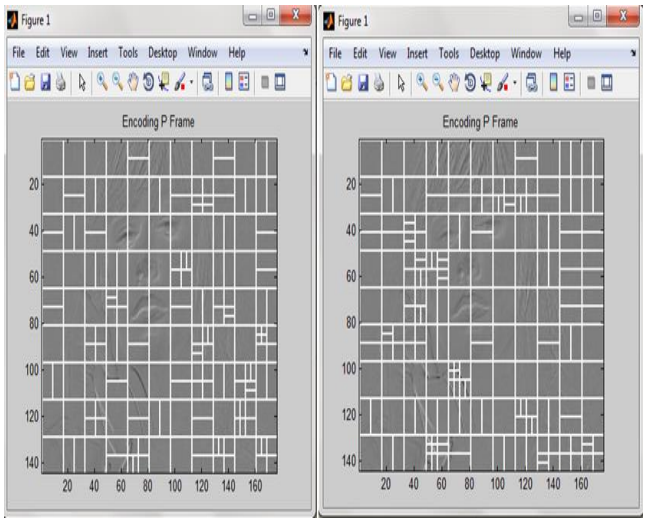


Figure 3 Proposed Block Diagram Of H.264

In figure 3 represents the encoder and decoder of the proposed techniques. It gives the detail explanation of the process of the video sequence uses the intra and inter frame formats in terms of compression efficiency at the time of writing (early 2010). The Moving Picture Experts Group (MPEG) and Video Coding Experts Group (VCEG) are examining the need for a new video compression standard. The consensus was that (a) there is likely to be a need for a new compression format, as consumers demand higher-quality video and as

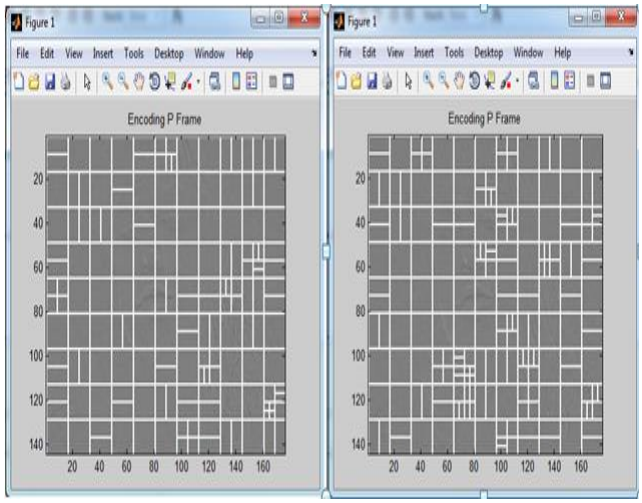
processing capacity improves and (b) there is potential to deliver better performance than the current state-of-the art. A number of different techniques were proposed, including decoder-side motion estimation, larger macro block sizes (up to 32 × 32), and more sophisticated in-loop Deblocking filters, adaptive transform sizes and improved intra prediction. In general, all of these proposed algorithms offer the potential for better compression performance at the expense of increased computational complexity. Results of subjective comparison tests of the proposed architecture to conclude that 'for a considerable number of test sequences significant peak signal noise ratio (PSNR) could be achieved'. prediction, later on motion Compensation and motion Estimation is done on the frames by using the reference frame.

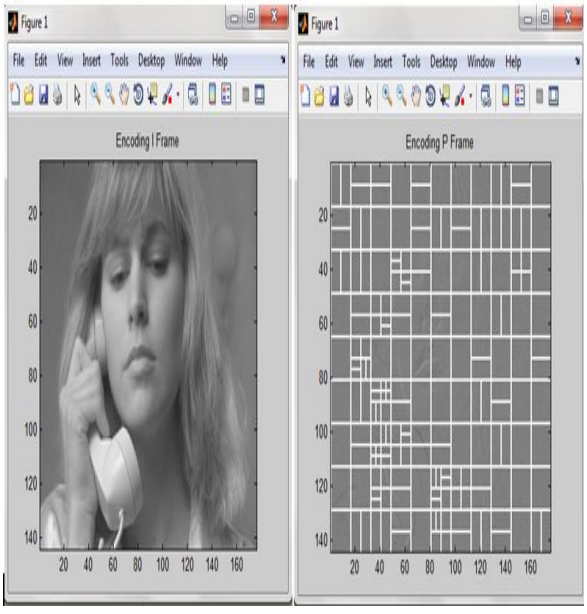


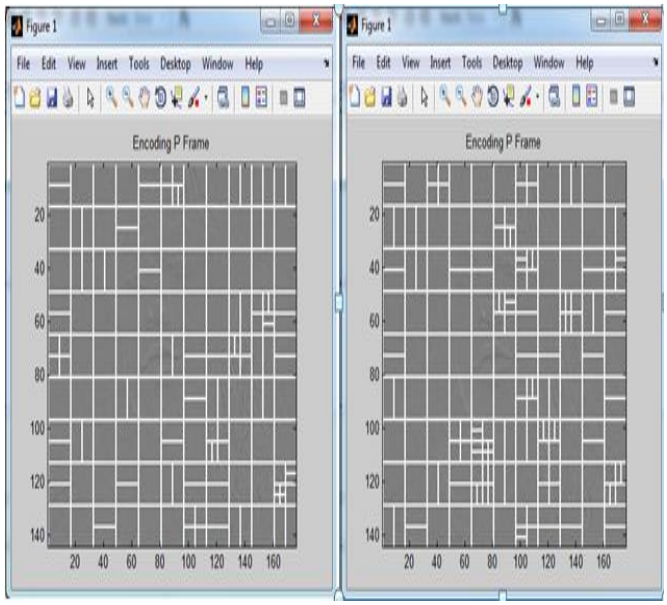
IV. SIMULATION RESULTS

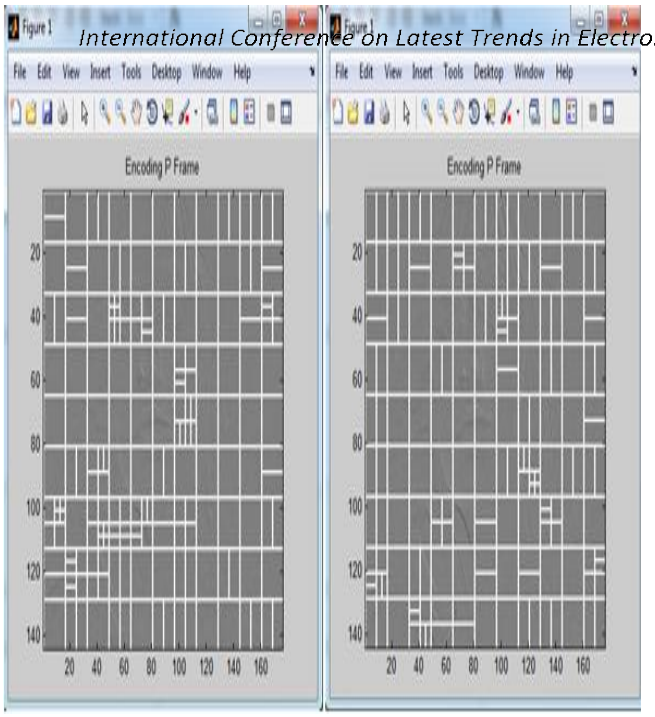
Encoding Process:

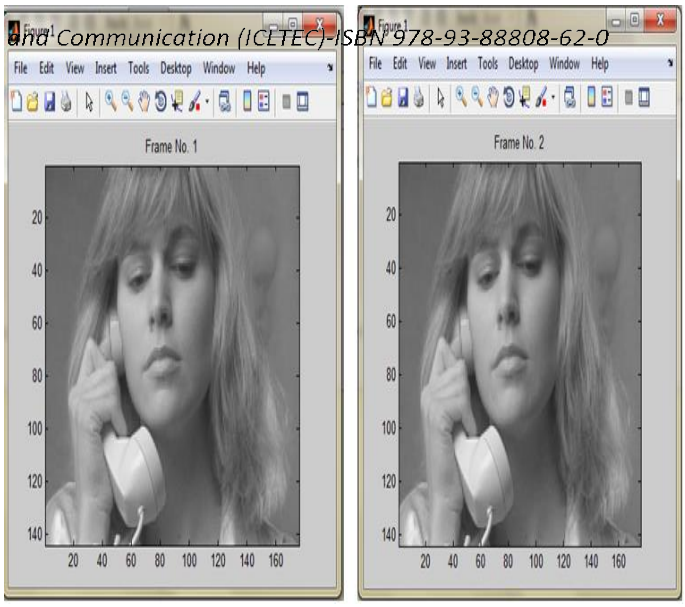
In this process 'suzie_qcif.yuv' is considering as a input video sequence. The block diagram of the basic encoder with extensions for non-intra frame coding techniques is given. Of course, this encoder can also support intra frame coding as a subset. Starting with an intra, or I frame, the encoder can forward predict a future frame. This is commonly referred to as a P frame, and it may also be predicted from other P frames, although only in a forward time manner. As an example, consider a group of pictures that lasts for 10 frames. In this case, the frame ordering is given as I, P, P, P, P, P,I,P,P,P,P,....sequence.

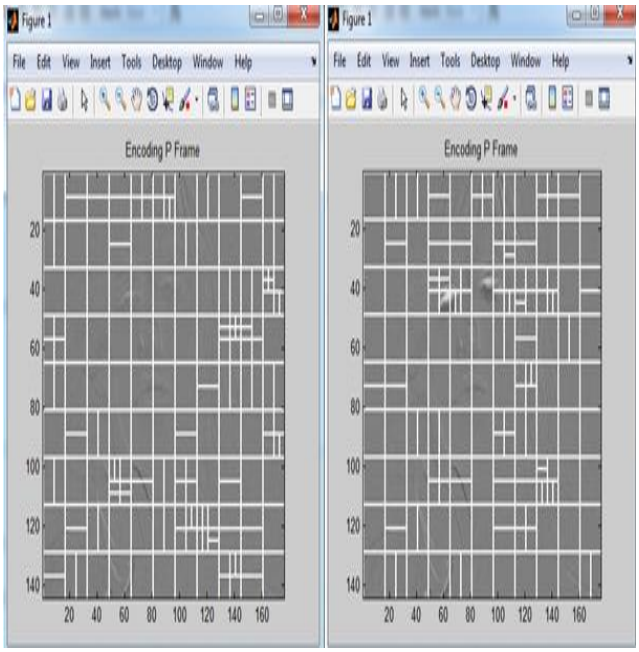


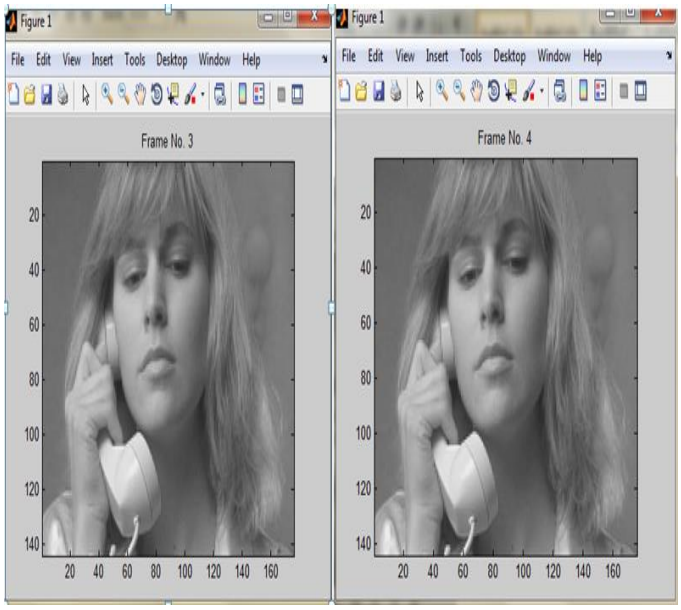


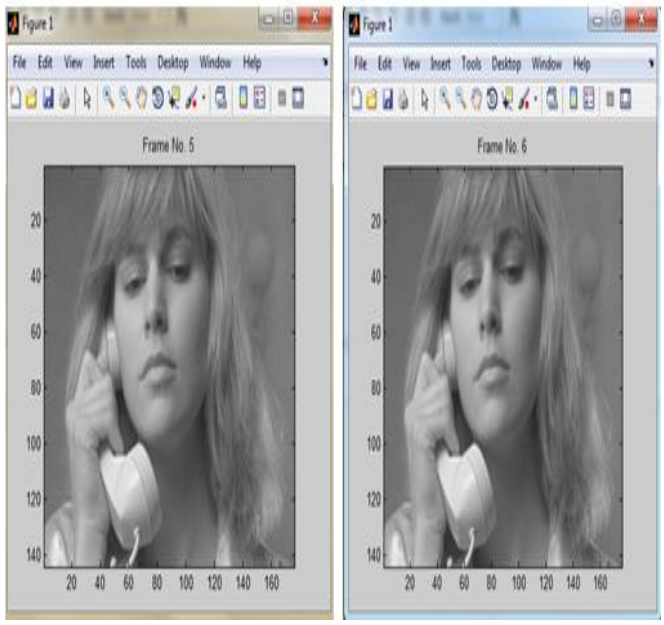






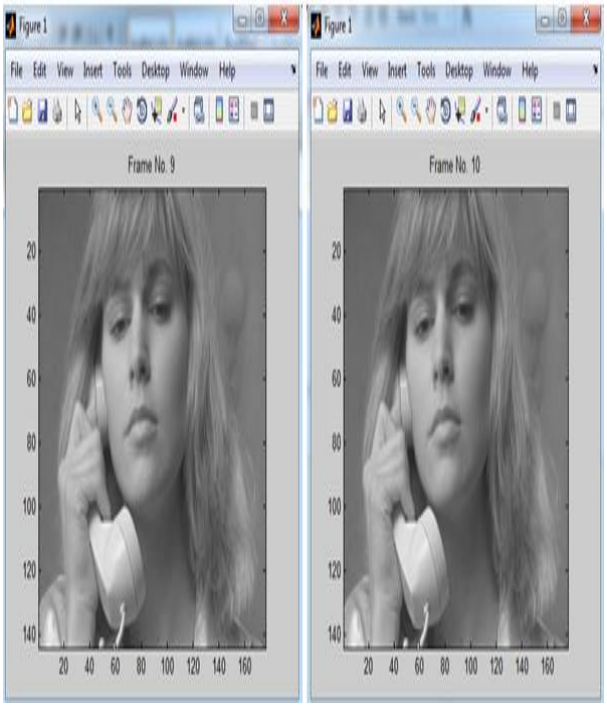






Decoding process:

In this process bit stream, The block diagram of the basic decoder with extensions for non-intra frame coding techniques is given. Of course, this decoder can also support intra frame coding as a subset. Starting with an intra, or I frame, the encoder can forward predict a future frame. This is commonly referred to as a P frame, and it may also be predicted from other P frames, although only in a forward time manner. As an example, consider a group of pictures that lasts for 10 frames. In this case, the frame ordering is given as Frame 1 , Frame 2 ,... Frame 10.



First standardized in 2003, H.264/AVC is now a relatively mature technology. H.264 is certainly one of the leading subjective quality of one point, his implies that there is scope for a new coding format that significantly out-performs H.264/AVC. The current plan is to set up a Joint Collaborative Team (JCT) of MPEG and VCEG representatives to work on a new video coding standard. In this work, considered 'suzie_qcif.yuv' as test sequence its size is 176 width as well as 144 height, frame frequency is 30 Hz (30 frames per sec). The GOP test sequence is IPPPPPP-----, in this process I frame is having complete information of the scene or content of the image, whereas following P sequences is completely using comparison with I frame with process of motion estimation and compensation accounted as motion prediction technique such as inter and intra prediction. The split procedure partitions the MB into variable size block using a quad tree approach. In this method a macro block is divided into quarters of equal area Then using similarities of motion vectors of adjacent blocks we will show how to merge the sub-blocks for quarter division Loading a file and then generate the bit stream, the regular encoding process is to be done with motion estimation with different quantization parameters (QP) value. The same in the reverse process can be done in the Decoder, encoding frames and decoding frames As shown in the below

CONCLUSION

H.264/AVC represents a major steps in the development of video coding standards, in terms of both coding efficiency enhancement and flexibility for effective use over a broad variety of network types and application domains. In this work survey has been done in terms all the technical features such as Transformation in terms of various methods, motion estimation and compensation in terms of Inter and Intra Prediction and also the final section is Deblocking filtering process. Among them is enhanced motion prediction capability, use of small block-size exact-match transform, adaptive in-loop de-blocking filter, and enhanced entropy coding methods. The H.264/AVC is highly flexible with all motion model and the very efficient performance as comparing with the existing methods (H.261/H.262/H.263) performance of H.264/AVC as described in the project.

REFERENCES

[1]. T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Sys-temsforVideoTechnology*,vol.13,no.7,pp.560–576,July2003.
Samitha T,Study of H.264/Avc Algorithm and It's Implementation In Matlab,jan 2014.
[2]. Thomas Stutz, A Survey of H.264 AVC/SVC Encryption. VOL. 22, NO. 3, MARCH 2012.
[3]. Mohammed A. Saleh, Coding Structure and Performance of High Efficiency Video Coding (HEVC) and H.264/AVC.
[4].I.E.Richardson,*H.264andMPEG-4VideoCompression*.Hoboken,NJ:Wiley, 2003.
[5].K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms,Advantages, Applications*. Boston, USA: Academic Press,1990.
[6]. Damián Ruíz, fast algorithm for h.264/avc intra prediction based on discrete wavelet transform.
[7]T.Elarabi, hybrid wavelet - dct intra prediction for h.264/avc interactive encoder.
[8]Pengyu Liu, A fast and novel intra and inter modes decision prediction algorithm for H.264/AVC based-on the characteristics of macro-block,2009
[9]Ashika K O, Inter Frame Coding in Advanced Video Coding Standard H.264 using Block Based Motion Compensation Technique,2017.
[10]M. Chan, Y. Yu, and A. Constantinides, "Variable size block matching motion compensation with applications to video coding," *IEEE Proceedings on Com- munication, Speech, and Vision*, vol. 2, no. 4, pp. 205–212, Aug.1990.
[11]Asim Munir, Deblocking Filters for Block-Based ICA Transform Coding,2010.
[12].Nikolaos Kefalas, a parallel luma-chroma filtering Architecture for h.264/avc deblocking filter,2015.

[13]Y. Lee and H. Park, "Loop filtering and post-filtering for low-bit-rates moving picturecoding,"*SignalProcessing:ImageCommunications*,vol .16,no.9,pp. 871–890, June2001.
[14]Tsu-Ming Liu, An In/Post-Loop Deblocking Filter With Hybrid Filtering Schedule,2007.
[15] Vinayak Ashok Bharadi, Hybrid Wavelet Transform I and II Combined with Contrast Limited Adaptive Histogram Equalization for Image Enhancement,2015.
[16] M. Krupa Swaroopa Rani, Implementation Of Image Compression Using Hybrid Dwt -Dct Algorithms.
[17]U. S. Mohammed and W. M. Abd-elhafiez, "Image coding scheme based on object extraction and hybrid transformation technique," Int. Journal of Engineering Science and Technology, vol. 2, no. 5, pp. 1375–1383, 2010.
[18]Suchitra Shrestha and Khan Wahid, "Hybrid DWT-DCT Algorithm for Biomedical Image and Video Compression Applications",2010.
[19].A. Alice Blessie1, J. Nalini and S.C.Ramesh, "Image Compression Using Wavelet Transform Based on the Lifting Scheme and its Implementation", May2011.
[20]Subhasis Saha, "Image compression - from DCT to Wavelets- AReview",2009.
[21]A. Puri, H. Chen, A. Luthra, "Video Coding using the H.264/MPEG-4AVC compression standard," *Signal Processing: Image Communication*, October 2004.
[22]Yu-Wen Huang, Analysis and Complexity Reduction of Multiple Reference Frames Motion Estimation in H.264/AVC,Apiril 2006.
[23]Jamil-ur-Rehman,Efficient Techniques for Signalling Intra Prediction Modes of H.264/Mpeg-4 Part 10,2006.
[24]Digital Image Processing, Rafael C. Gonzalez ,2009.

Low Power Viterbi Decoder Design based on Reversible Logic Gates

Nenavath Naresh

PG Student, Dept. of ECE, MRCE, Hyderabad, India, Email: nareshnenavath4@gmail.com

***Abstract*— In recent trends of VLSI technology the reversible logic has became the major area of research in optimization of area, power and speed constraints. The reversible logic has equal number of inputs and outputs. In wireless communications Viterbi algorithm is employed to have minimal number of communication channels. The Viterbi decoder design at 65nm technology using reversible logic has made an attempt for optimizing power, area and delay with increased efficiency.**

***Keywords*—** *Viterbi decoder, Reversible logic, Feynman Gate, Peres Gate, HN Gate, Power consumption, Delay.*

I. INTRODUCTION

The Viterbi decoding algorithm was proposed by Andrew J Viterbi, which is a decoding process for convolutional codes in memory-less noise. This algorithm is implemented in the designing of communication systems. The Viterbi Algorithm is the most resource consuming and it finds the most-likely noiseless state transition sequence in a state diagram, given a sequence of symbols which are interrupted by noise [1].

Generally, a viterbi decoder consists of three basic computation units: Branch Metric Unit (BMU), Add-Compare-Select Unit (ACSU) and Survivor Memory Unit (SMU) [2].

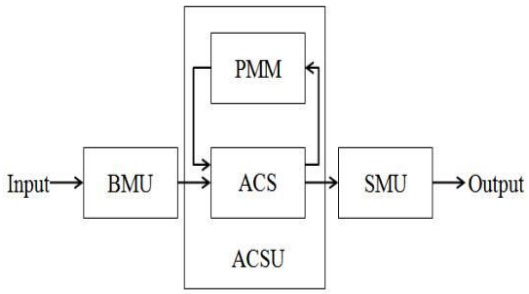


Figure 1. Block diagram of Viterbi decoder

The primary unit is called as Branch Metric Unit (BMU). This unit will compare the received data symbols with the ideal outputs of the encoder and finally the branch metric will be calculated. The Euclidean distance or Hamming distance is utilized for the calculation of branch metric. The BMU creates branch metrics for the following module in terms of the symbols which are received by the channel.

The Add Compare Select Unit (ACSU) is illustrated as the sum of the Branch Metrics (BM) to the subsequent Path Metrics (PM). The new PM will be compared and the selected PM will be stored in the Path Metric Memory (PMM). At the same time, the ACSU stores the associated survivor path decisions in the Survivor Memory Unit (SMU).The PM of the survivor path of each state is updated and stored back into the PMM.

The Survivor Memory Unit uses the Trace-Back method to identify the survivor path and output data. In this unit the decoded bits are extracted from the beginning through

Minimum path metric. At the beginning state, backward tracing is followed by the survivor path, which initially contributed to the current PM and a unique path is identified. While tracing back through the trellis, the decoded output sequence corresponding to the traced branches is generated in the reverse order.

II. REVERSIBLE LOGIC

Reversible computing is the application of principles of recycling to computing. A reversible logic gate is mapped with one-to-one logic device having an n-input, n-output gate. As it helps to find out the outputs from inputs although the inputs can be exclusively recovered from the outputs. In the necessary conditions to have the number of inputs equal to the number of outputs additional inputs or outputs is added. An important constraint present on the design of a reversible logic circuit using reversible logic gate is that the fan-out is not allowed. The quantum cost of reversible logic circuits must be minimum. With the minimum number of reversible gates the design of reversible circuit is accomplished. The major constraint to attain optimization of the circuit is to produce the garbage outputs and the constant inputs with the minimum number.

The reversible logic gates are the circuits which has number of inputs is equal to number of outputs. The important optimization parameter for every reversible logic gate is the quantum cost [3].

The important reversible logic gates which are required for designing Viterbi decoder are.

A. Feynman Gate

The Feynman gate is 2x2 reversible gate which has the inputs (A, B) and the outputs (P =A, Q = A⊕ B). This gate is also referred as Controlled NOT. The quantum cost is 1. This is mainly used for the fan-out function. The power consumption and delay are 18mW and 7.760ns [4].

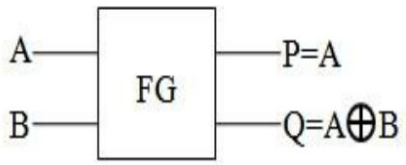


Figure 2. Feynman Gate

B. Peres Gate

The Peres gate is 3x3 reversible gate, where the inputs are A, B, C and the outputs are P, Q and R. The outputs are mapped as P =A, Q = A⊕ B and R= A.B⊕ C. The quantum cost is 4. The power consumption and delay are 24mW and 7.824ns.

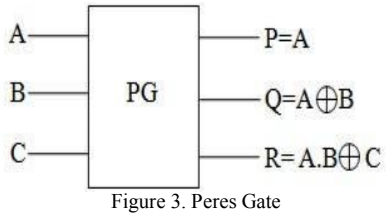


Figure 3. Peres Gate

A. HN Gate

The HNG is a 4x4 reversible gate, and abbreviated as Haghparast Navi Gate (HNG). It has 4 inputs A, B, C, D and 4 outputs P, Q, R, S, its mapping is P=A, Q= B, R= A⊕ B⊕ C and S= (A ⊕ B) C ⊕ AB ⊕ D. The quantum cost is 6. The power consumption and delay are 24mW and 7.824ns.

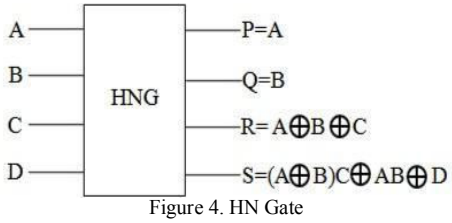


Figure 4. HN Gate

III. PROPOSED VITERBI DECODER USING REVERSIBLE LOGIC GATES

In proposed system the calculation of metric, addition, comparison of weight and selection of survivor path everything is carried out in the ACS array (ACSU). Thus the ACS array contains the values of weight at every state, as a progression along the survivor path. The codewords are very much required when comparing with different codewords of different paths, to identify the survivor path. The functions of compilation and comparison takes place within the ACS array which actually identifies the extensions of path. The signals determining these extensions to the survivor paths are passed from the ACS array to the SPU, which then updates the survivor paths.

The Figure 5 shows the top level RTL design of Viterbi decoder in which U1 indicates compute metric unit, U2 represents metric unit, U3- acs enable unit, U4- compare select unit, U5- reduce unit and U6- path memory unit.

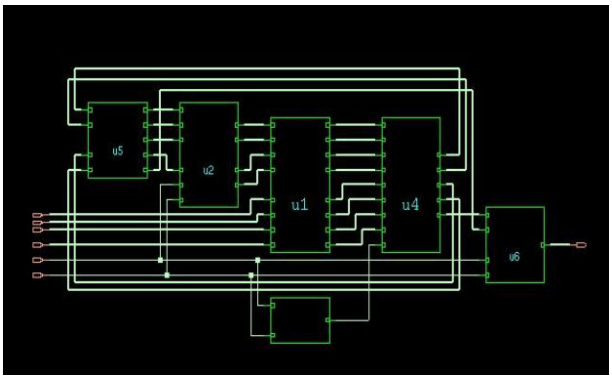


Figure 5. Top Level RTL design of Viterbi Decoder using Reversible Logic Gates

The design has the four input signals each signal having two bits with clock and reset signals. As the Reversible Logic gates reduces the power consumption and path delay, the PG will function as half adder and HNG will function as full adder.

A. Compute Metric Unit

In the compute metric unit, each received code word and each excepted code word (metric output) will be the inputs for this unit in which if performs the comparison between the code words and generates a codeword which is the input for the compare select unit.

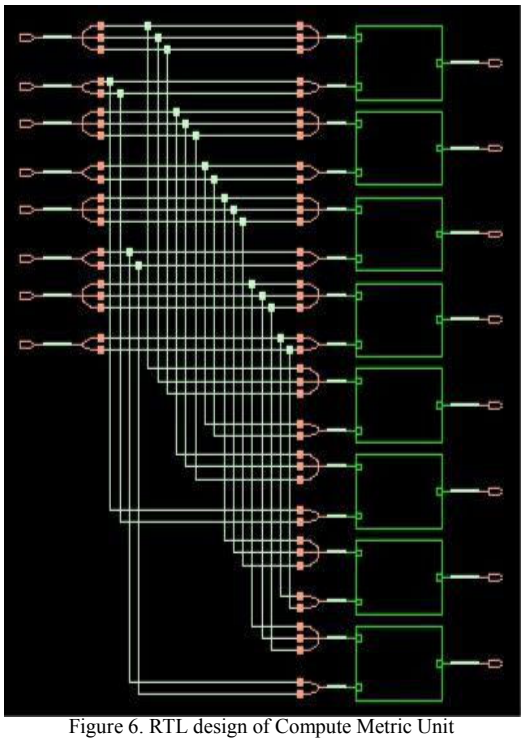


Figure 6. RTL design of Compute Metric Unit

The Figure.7 shows a unit of compute metric in which expected bits are represented as a [0:2] and received bits as b [0:1]. The blocks U_FA_0 and U_FA_2 is implemented by Peres gate which performs the half adder and U_FA_1 is implemented by HNG performing full adder.

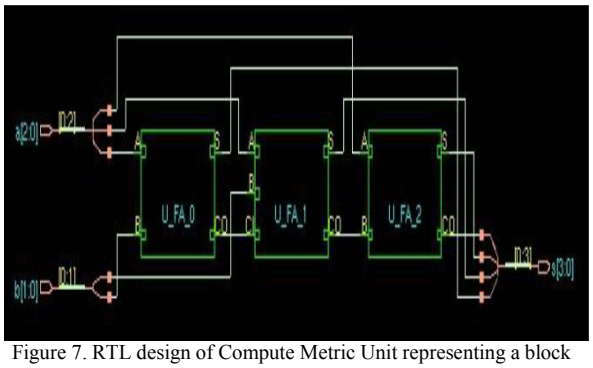


Figure 7. RTL design of Compute Metric Unit representing a block

Each bit of the metric unit will be the inputs for the 3 sub units and user input of 2 bits will be for the inputs for U_FA_0 and U_ FA_ 1. The U_FA_0 will generates the output of 2 bits, one bit will be the final output of the block and other bit will be the input for the next sub unit i.e. U_FA_1. The U_ FA_1 will generates the output of 2 bits, one bit will be the final output of the block and other bit will be the input for the next sub unit i.e. U_FA_2. Similar function is performed by the U_FA_2 sub unit and both the outputs. The unit has 8 blocks and each block performs the same function as explained above. Hence the compute metric has 8 outputs with each output of 4 bits.

B. Metric Unit

The metric unit performs the storage operation of the expected code word. The Figure 8 shows the RTL design in which it has the cock and reset signals for each metric unit.

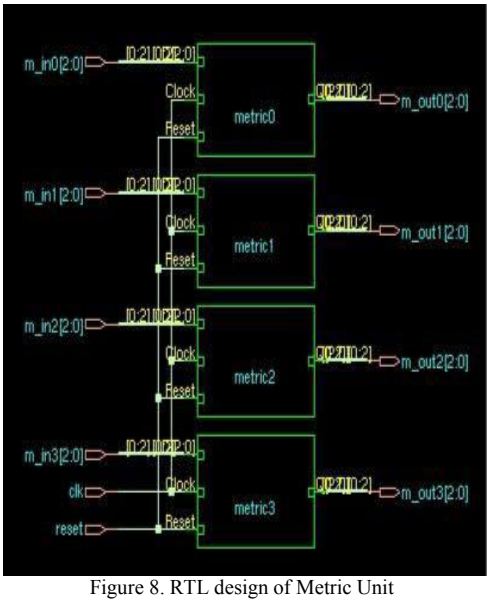


Figure 8. RTL design of Metric Unit

The metric unit stores the minimum metric of the oldest bits as it is required for next computation cycle. Every time the oldest path will be erased and the new minimum reduced metric will be updated. The unit has 4 metric blocks and each metric block has 3 inputs and 3outputs.

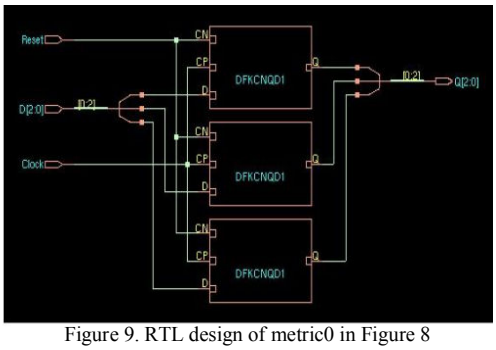


Figure 9. RTL design of metric0 in Figure 8

Figure 9 represents the each metric unit which consists of three D flip-flops, once the clock signal is applied upon reset=1 the bits will perform the comparison operation in the compute metric unit.

C. ACS Enable Unit

Figure 10 shows the acs-enable unit which consists of D flip-flops which has clock and reset signals and initially the reset will be 0, D input is 0.

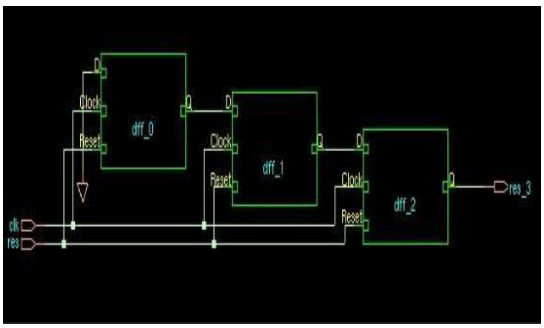


Figure 10. RTL design of ACS-Enable Unit

Once the reset value becomes 1 the unit will send a signal such that a compare-select unit is enabled.

D. Compare Select Unit

The codeword which is received from the compute metric unit will have maximum path metric such that the minimum path metric has to be calculated by selecting the appropriate metric, this operation in performed by the compare select unit and Figure 11 shows the RTL design of the unit.

Upon receiving the acs enable signal this unit will be enabled and the each received codeword from the compute metric unit will be of 4 bits. Two code words will be added and compared such that minimum metric will be calculated and obtained by performing the logical operations such as AND and OR. Finally the minimum metric will be selected and the codeword will be three bits. The output will be of four bits which will be stored in the path memory. The minimum metric which is calculated will be reduced such that previous path will be deleted. The path metric of both inputs will be compared and the metric with smallest distance will be the output and this is considered as the control signal for the path memory unit.

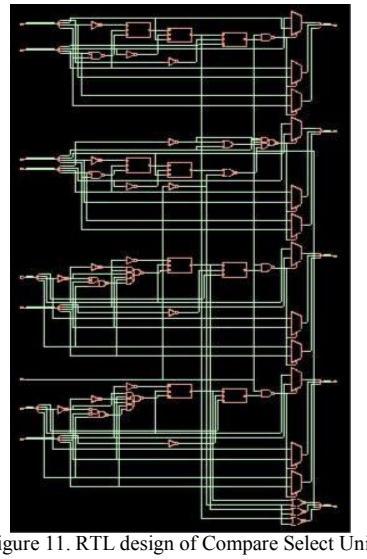


Figure 11. RTL design of Compare Select Unit

E. Reduce Unit

The minimum path metric which is obtained has to compared with the oldest path metric and it will be stored in the reduce unit, suppose if both the metric are equal then the oldest path metric will be retained otherwise the new minimum path metric will be stored in the reduce unit.

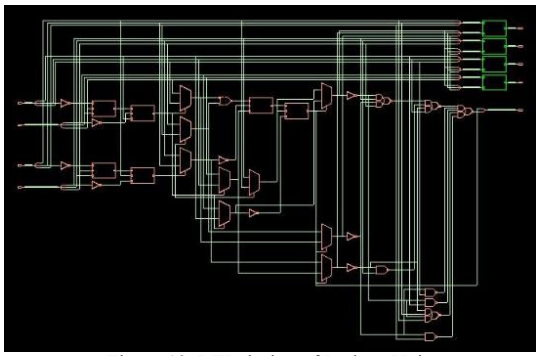


Figure 12. RTL design of Reduce Unit

The old minimum path metric which is stored will be evaluated with the HN and Feynman reversible gates such that the new minimum metric of 3 bits will be stored by eliminating the oldest path.

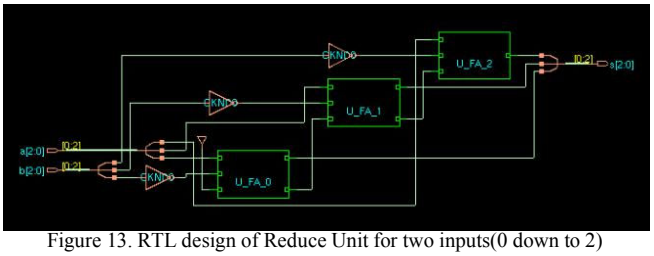


Figure 13. RTL design of Reduce Unit for two inputs(0 down to 2)

The control signal of 2 bits is required for the path memory to enable the unit and this is generated by the reduce unit by latch which consists of OR and AND logical gates.

F. Path Memory Unit

The path memory unit consists of 4x1 multiplexer units and the D flip -flops. The control signal is considered as a select signal for the multiplexer as it is used to select the minimum path metric of the decoded message bit.

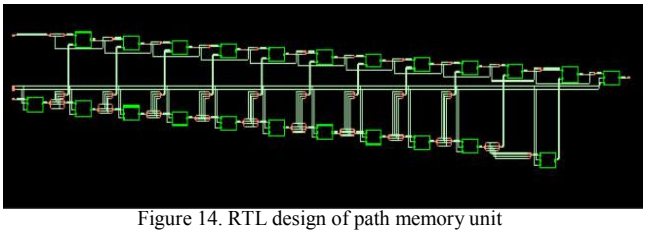


Figure 14. RTL design of path memory unit

The input for the D- flip flop is the ACS signal which is the output from the compare-select unit of 4 bits. These bits are considered as a survival bit inputs for the 4x1 multiplexer.

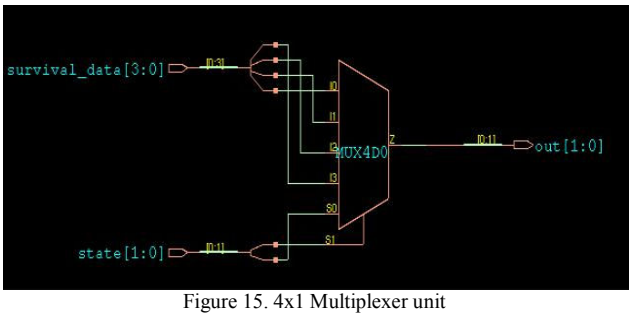


Figure 15. 4x1 Multiplexer unit

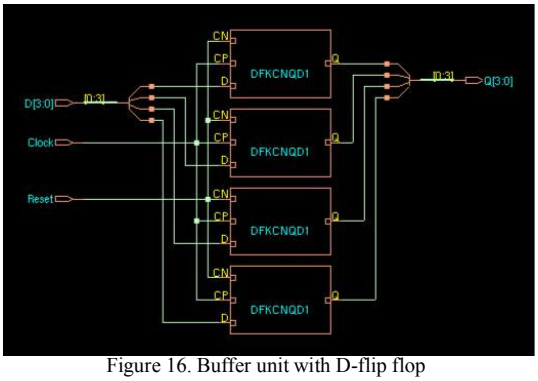


Figure 16. Buffer unit with D-flip flop

During the initial stage the path 1 is identified, then path 2 is identified by the multiplexer units. The unit consists of 11 multiplexer units and 9 D flip-flops.

The decoder output will be available at the receiver end after the completion of 14 clock cycles. Since the clock and reset signal are applied to the acs enable and path memory units.

IV. Results and Discussion

The proposed Viterbi decoder is developed by Verilog coding and simulated. The main constraints for VLSI design is area, speed and power. The power performance report is shown in Table 1, area and timing performance report is tabulated in Table 2.

Table 1. Power Performance Summary

INSTANCE	Power Analysis in ηW		
	Leakage Power	Dynamic Power	Total Power
U1 [Compute Metric]	1436.774	4871.918	6308.693
U2[Metric]	555.193	2961.352	3516.545
U3[ACS_Enable]	157.896	298.383	456.279
U4[Compute Select]	520.295	3986.532	4506.828
U5[Reduce]	1000.059	5939.455	6939.514
U6[Path Memory]	2697.978	11326.026	14024.004
VITERBI [TOP LEVEL]	**6368.196**	**30094.526**	**36462.722**

Table 2. Area and Timing Performance Summary

INSTANCE	Number of Cells	Area (sq microns)	Time Delay (psec)
U1 [Compute Metric]	40	173	320
U2[Metric]	12	86	135
U3[ACS_Enable]	3	22	0
U4[Compute Select]	51	111	533
U5[Reduce]	74	177	1343
U6[Path Memory]	57	422	2213
VITERBI [TOP LEVEL]	**237**	**990**	**4544**

The functionality of Viterbi decoder using reversible logic gates is verified and simulated. The simulation waveform result is shown in Figure 17.

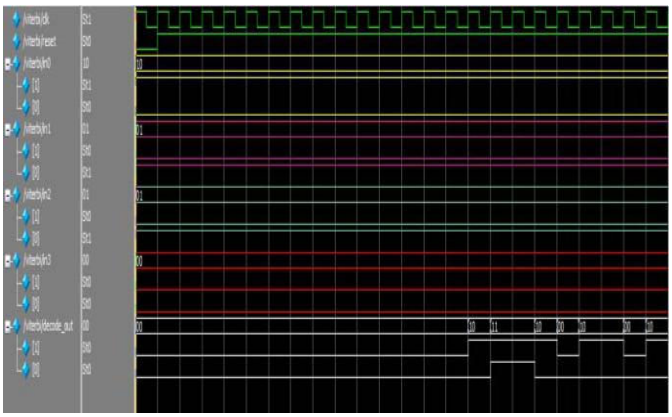


Figure 17. Schematic waveforms

Clk & reset; in0 (input); in1 (input); in2 (input); in3 (input); decode_out

V. CONCLUSION

The Viterbi decoder which consists BM, ACS and SM Units have been implemented by considering the reversible logic gates such as Feynman, Peres and HN gates to achieve a better efficiency. The designing is done using HDL coding and synthesized in CADENCE tool. The performance summary such as area, power efficiency and delay results are achieved. Hence this design can be used in high speed communication applications.

REFERENCES

[1] Abdulfattah M. Obeid, Alberto Garcia, "*A Multipath High Speed Viterbi Decoder*", IEEE 2003: pp.1160-1163.

[2] Dalia A. El-Dib and Mohamed I. Elmasry, "*Modified Register-Exchange Viterbi Decoder for Low-Power Wireless Communications*", IEEE Transactions on Circuits and Systems, Vol.51, No.2 February 2004: pp.371-378.

[3] H.R.Bhagyalakshmi, *"An Improved Design of a Multiplier Using Reversible Logic Gates",* International Journal of Engineering Science and Technology, Vol. 2(8), 2010, 3838-3845.

[4] B. Murali Krishna, B. Raghu Kanth, " *A Comparative Study of Reversible Logic Gates",* International Journal of VLSI and Signal Processing Appications, Vol 2, Issue 1, Feb 2012, pp 51-55.

[5] Alexander Fish, Arkadiy Morgenshtein "*Gate-Diffusion Input(GDI) – A novel Power Efficient Method for Digital Circuits*", IEEE 2001: pp. 39-43.

[6] Li Qiao, You Yuxin, Wang Jinxiang and Ye Yizheng, "*VLSI Implementation of a High Speed and Low Power punctured Viterbi Decoder*", International Conference on Computers, Communications, Control and Power Engineering, IEEE transactions ,October 2002.

[7] C.Arun, V Rajamani, " *Design and VLSI Implementation of a Low Probablity of Error Viterbi Decoder",* First International Conference on Emerging Trends in Engineering and Technology, ICETET, July 2008.

[8] D. Chakraborty, P. Raha, A. Bhattacharya, *"Speed Optimization of a FPGA Based Modified Viterbi Decoder",* International Conference on Computer Communication and Informatics (ICCCI), January 2013.

[9] Landauer.R, "*Irreversibility and Heat Generation in the Computing Process*", IBM J. Research and Development: pp. 183-191, 1961.

[10] Song Li, Qing-Ming Yi, "*The Design of High-Speed and Low Power Consumption Bidirection Viterbi Decoder",* International Conference on Machine Learning and Cybernetics, August 2006.

[11] Hao Yang and Xiao-Lang Yan," *Design and Implementation of a High-Speed and Area-Efficient Viterbi Decoder",* 8th International Conference on Solid-State and Integrated Circuit Technology, ICSICT October 2006.

[12] Umar M Karim, Muhammad U.K Khan, Yahya M. Khawaja, "*An Area Reduced, Speed Optimized Implementation on Viterbi Decoder",* International Conference on Computer Networks and Information Technology (ICCNIT), July 2011.

[13] R.Tessier, S. Swaminathan, R. Rangaswamy, "*A Reconfigurable, Power-Efficient AdaptiveViterbi Decoeder",* IEEE Transactions on Very Large Scale Integration Systems, April 2005.

[14] F.Chan and D.Haccoun, *" Adaptive Viterbi Decoding of Convolutional Codes Over Memoryless Channels",* IEEE Transactions on Communications, 45(11): 1389-1400, November 1997.

[15] Kai-Ting Shr, Hong-Du Chen, Yuan-Hao Huang, " *A Low Complexity Viterbi Decoder for Space-Time Trellis Codes"*, IEEE Transactions on Circuits and Systems, April 2010.

[16] Chun-Yuan Chu, Yu-Chuan Huang, An-Yeu Wu," *Power Efficient Low Latency Survivor Memory Architecture for Viterbi Decoder",* IEEE Transactions 2008.

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

DESIGN OF HIGH SPEED SEQUENCE DETECTOR USING VERILOG

Bejagam Divya[1]
M.Tech Scholar[1]
Department of Electronics and Communication Engineering[1]
Malla Reddy College of Engineering, Hyderabad, India[1]
bejagamdivya84@gmail.com[1]

Abstract: There is a enormous usage of sequence detectors in digital circuits as it is the basic function and it became essential in most of the digital systems counting ALU, microprocessors and DSP. Sequential circuit's works on a clock cycle which may be synchronous or asynchronous. Sequential circuits use current inputs and previous inputs by storing the information and putting back into the circuit on the next clock cycle. This paper presents the high speed Sequence Detector in Verilog, which is a sequential state machine used to detect consecutive bits in a binary string. The flip-flops help to detect the pattern in the given string. The Sequence Detector gives for some particular sequence of inputs and outputs, whenever the desired sequence has found. And this paper shows a great vision on the design analysis of sequence detector using Verilog. The delay (1.045ns) minimized. The proposed architecture of sequence detector is synthesized in Xilinx ISE14.7.

Keywords: Sequence detector, sequential circuits, flip-flop, delay, Verilog, Mealy machine and Xilinx ISE14.7.

I. Introduction

FSM is a mathematical model of computation. It is an abstract machine that can be in exactly one of a finite number of states at any given time. The FSM can change from one state to another in response to some external inputs; the change from one state to another is called a transition. An FSM is defined by a list of its states, its initial state, and the conditions for each transition. Finite state machines are of two types - deterministic finite state machines and non-deterministic finite state machines. A deterministic finite-state machine can be constructed equivalent to any non-deterministic one.

The behavior of state machines can be observed in many devices in modern society that perform a predetermined sequence of actions depending on a sequence of events with which they are presented. Simple examples are vending machines, which dispense products when the proper combination of coins is deposited, elevators, whose sequence of stops is determined by the floors requested by riders, traffic lights, which change sequence when cars are waiting, and combination locks, which require the input of combination numbers in the proper order. The finite state machine has less computational power than some other models of computation such as the Turing machine.

II. Proposed work

A sequence detector is a sequential state machine. Sequential circuit's works with respect to a clock cycle which may be synchronous or asynchronous. The figure shows a basic diagram block of sequence detector. Sequential circuits use current inputs and previous inputs by storing the data and putting back into the circuit on the next clock cycle.

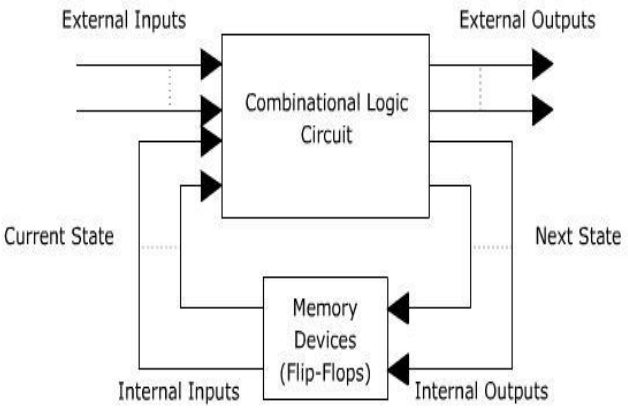


Figure 1: Block Diagram of Sequence Detector.

Finite State Machine (FSM):

A FSM is a model it is used to design sequential logic circuits. It is considered as an abstract machine that can be in one of a finite number of states. The machine is accessible in only one state at a time; the state it is in at any given time is called the current state. It can change

from one state to another state when the triggering event or condition is introduced into the machine, this is called a transition. A specific FSM is well-defined by a list of its states, and the triggering condition for each transition. It can be implemented using models like Mealy machine and Moore machine. For this expt., we will use Mealy machine model implementation.

Types Of Sequence Detector:
There exist basically two types of sequence detectors.
That are:
1) Overlapping Sequence Detector.
2) Non- Overlapping Sequence Detector.

Overlapping Sequence Detector:
In a sequence detector that allows overlap, the final bits of one sequence can be the start of another sequence.

Non-Overlapping Sequence Detector:
The sequence detector with no overlap allowed resets itself to the start state when the sequence has been detected.

Flip-flop: A flip-flop or latch is a circuit that has two stable states and it can be used to store state data. A flip-flop is a bistable multivibrator. The circuit can be made to change state by signals applied to one or more Control inputs and will have one or two outputs, one for the normal value and one for the complement value of the stored bit. Memory elements in any sequential circuit are usually flip-flops.

Sequence detector:
Suppose a sequence detector is to be designed to detect a sequence 1101.
Then the state diagram will be:

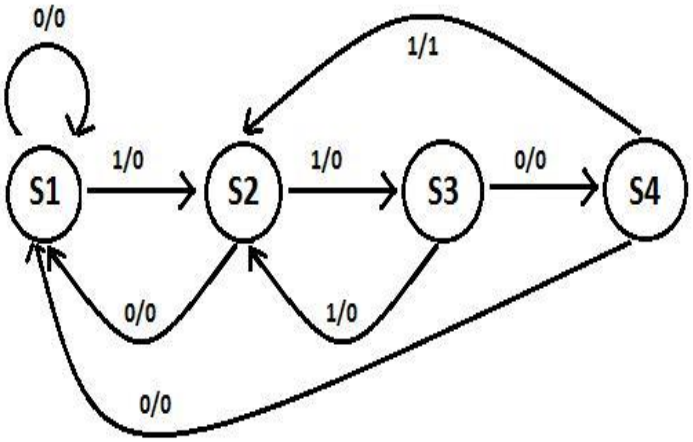


Figure 2: State Diagram of Sequence Detector.

Note that this state diagram is considering overlap i.e. if we have input 1101101 we will have output0001001. For same input, non-overlap case will have output 0001000. Either cases are correct but we will consider only overlap case henceforth.

III. Synthesis Results

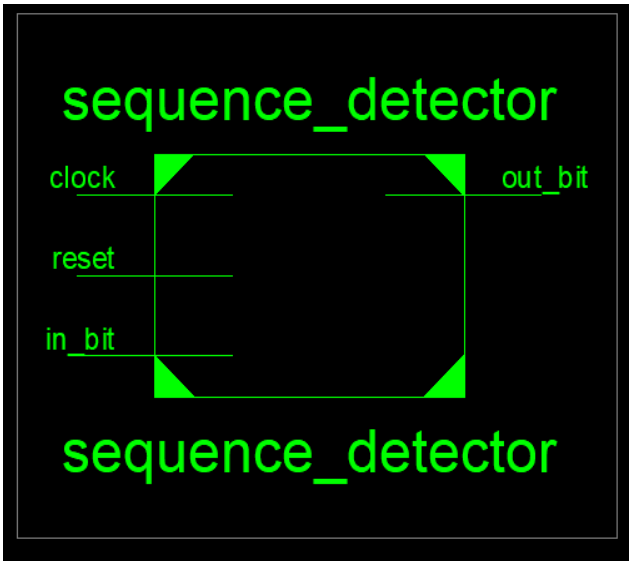


Figure 3 : Top Module of Sequence Detector.

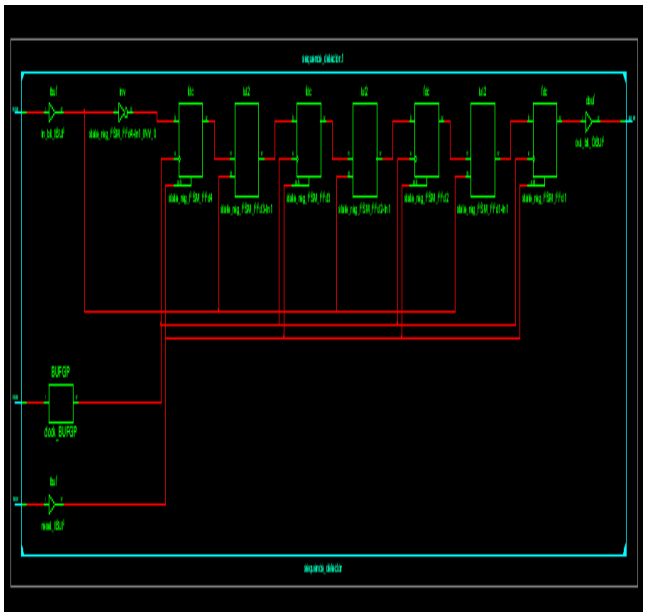


Figure 4 : Technology Schematic of Sequence Detector.

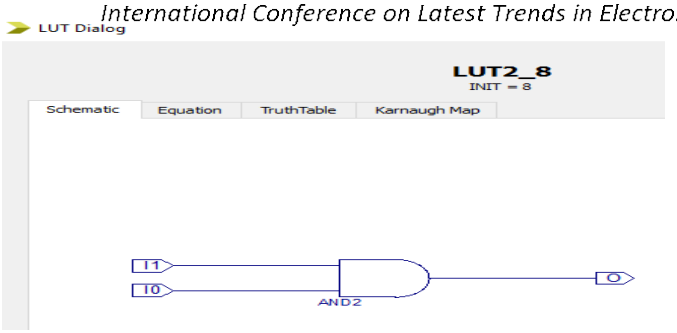


Figure 5 : LUT Schematic of Sequence Detector.

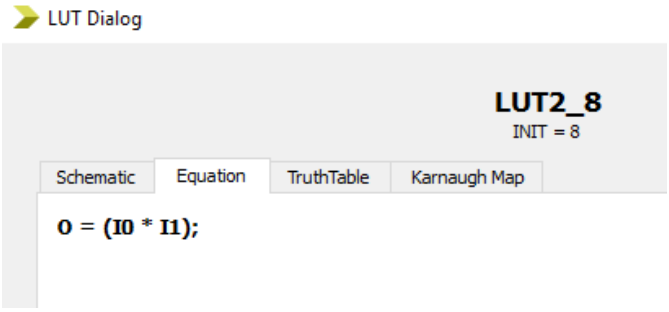


Figure 6 : LUT Equation of Sequence Detector.

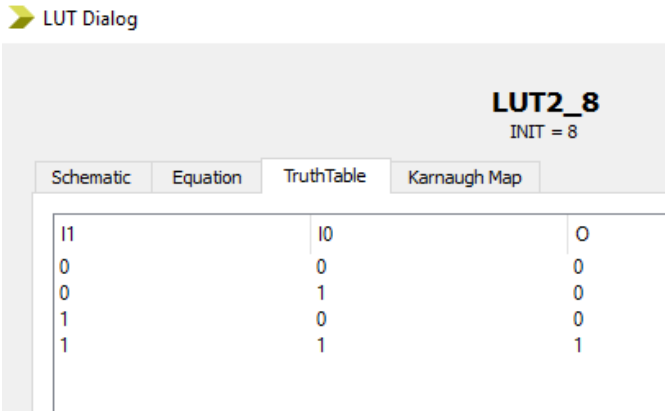


Figure 7 : LUT Truth table of Sequence Detector.

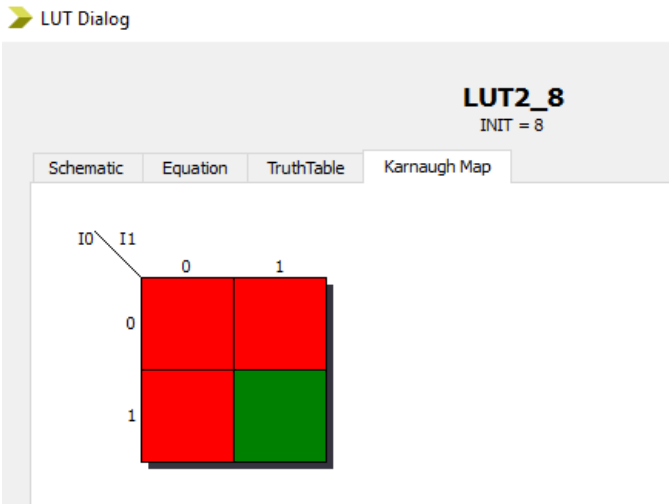


Figure 8 : LUT karnaugh map of Sequence Detector.

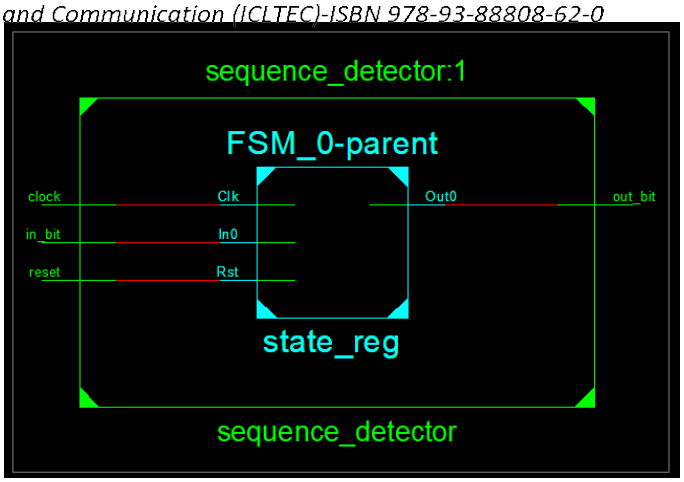


Figure 9 : Top RTL Schematic of Sequence Detector.

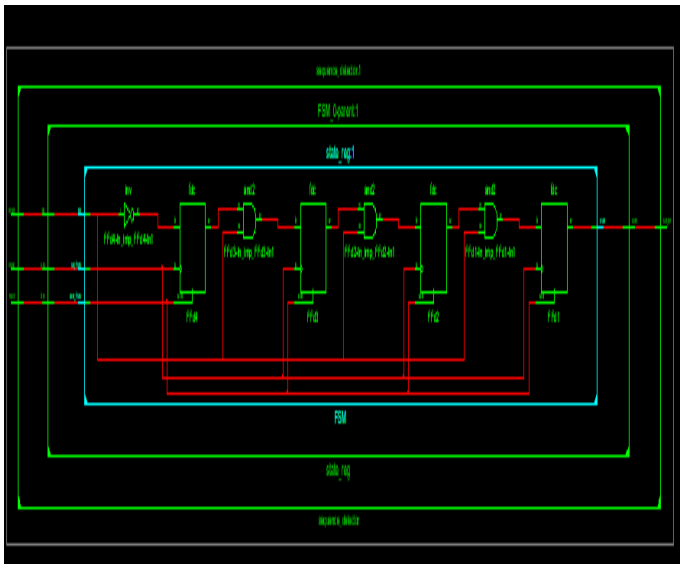


Figure 10 : RTL Schematic of Sequence Detector.

Device Utilization Summary (estimated values)			[-]
Logic Utilization	Used	Available	Utilization
Number of Slice Registers	4	44800	0%
Number of Slice LUTs	4	44800	0%
Number of fully used LUT-FF pairs	0	8	0%
Number of bonded IOBs	4	640	0%
Number of BUFG/BUFGCTRLs	1	32	3%

Table 1: Device Utilization of Sequence Detector.

Final Register Report:

Macro Statistics

# Registers	: 4
Flip-Flops	: 4

Cell Usage :

# BELS	: 4
# INV	: 1
# LUT2	: 3
# FlipFlops/Latches	: 4
# FDC	: 4
# Clock Buffers	: 1
# BUFGP	: 1
# IO Buffers	: 3
# IBUF	: 2
# OBUF	: 1

Timing Summary:

Speed Grade: -1

Minimum period: 1.045ns (Maximum Frequency: 956.938MHz)

Minimum input arrival time before clock: 1.744ns

Maximum output required time after clock: 3.259ns

Timing Detail:

Timing constraint: Default period analysis for Clock 'clock'

Clock period: 1.045ns (frequency: 956.938MHz)

Total number of paths / destination ports: 3 / 3

Delay: 1.045ns (Levels of Logic = 1)

Source: state_reg_FSM_FFd2 (FF)

Destination: state_reg_FSM_FFd1 (FF)

Source Clock: clock rising

Destination Clock: clock rising

Total 1.045ns (0.565ns logic, 0.480ns route)
(54.1% logic, 45.9% route)

Total memory usage is **4568248 kilobytes**

IV. Simulation result

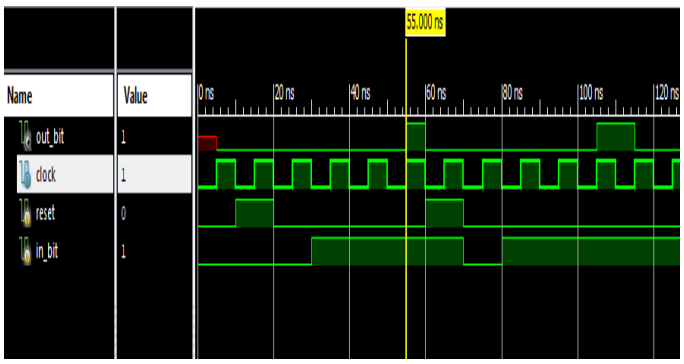


Figure 11: Simulation Result of the Sequence Detector.

The results are obtained by simulating the verilog code in Xilinx ISE 14.7.

V. Conclusion

The proposed design of sequence detector is synthesized and simulated in Xilinx ISE 14.7. The source code is written in Verilog. As we know Delay is the major factor in VLSI design that limits the performance of any circuit. This paper concentrates more on speed by presenting a simple approach to reduce the delay of sequence detector architecture, which helps in increasing the computational level of calculations. This proposed sequence detector has delay 1.045ns. Sequence detector has extensive variety of applications such as design of Ring counter, Serial Adder, Schmitt trigger.

References

1. IEEE Standard 1364-2001, IEEE Standard Description Language based on the Verilog Hardware Description Language, 2001.
2. Mindek, M., “Finite State Automata and Image Recognition” DATESO 2004, pp 132-143 (2004), ISBN: 80-248-0457-3.
3. G. Navarro, R. Baeza-Yates, “Improving an Algorithm for Approximate String Matching.”,Algorithmica, 30(4) 2001.
4. S.V.Ramasubramanian and Kamala Krithivasan,“Finite Automata and Digital Images”, IJPRAI, Vol. 14, No. 4, pp. 501-524, 2000.
5. John C.Martin , ”Introduction to Languages and Theory of Computation”,Third edition, Tata McGraw Hill,2009.

A low power large bandwidth flipped voltage follower based current mirror

P.Anil kumar1 **Rajesh Durgam2**
***1anilpole403@gmail.com , *2 rajeshdurgam2000@gmail.com**
1,2 Assistant Professor,Department of ECE, Malla Reddy College of Engineering, Hyderabad, Telangana, INDIA

Abstract
In this study, simple nMOS current mirror has been modified and the performance has been improved employing high performance flipped voltage follower (FVF). The resistively compensated FVF has been utilized at the input side of the current mirror. The advantages of using flipped voltage follower include high linearity, wide bandwidth and reduced power dissipation.

In this study, the simulation responses of all the circuits are presented. The functionality and performance improvement of all the circuits are simulated on Spectre simulator (Cadence) using model parameters of TSMC 0.18 μm CMOS BSIM 3 and level 49 technology.

Keywords: Current mirror, low power, analog circuit, cascode, voltage follower.

Introduction

The explosive growth in electronics world towards portability and high-speed VLSI systems has motivated the current research in the direction of high frequency analog basic cells. The requirement of high-performance analog devices in the communication systems has increased the usage of current-mode circuits. Current mirror is among the most important and essential current mode device that has been used in numerous analog systems.[1-9]

The important features of a current mirror include accurate current mirroring, large input and output current swings, high output impedance and good linearity. In the design of a current mirror the key issues are the improvement of high-frequency characteristic and the realization of high output impedance. Also, a current mirror which has high power consumption and small -3 dB frequency is not reliable for low-voltage high-speed applications. Several large bandwidth current mirrors are reported in literature.[1]

Gupta et al[10] have improved the frequency performance of FVF based current mirror by introducing both passive[3,4] and active resistance[3,4] at the gate of input transistor pair of the low voltage current mirror in. An approximately 200 MHz improvement in the bandwidth of passively and actively compensated current mirrors has been obtained.[10] In this work, the conventional current mirrors (CMs) have been redesigned for high frequency applications.

Simple Current Mirror

Voo and Toumazou[3] have improved the bandwidth of the simple current mirrors by using resistive compensation technique. They have introduced a compensating resistor in between the drain and gate of the input MOS transistor of the current mirror. This leads to introduction of one-pole and one-zero in the transfer function and the zero is used to cancel the dominant pole (pole-zero cancellation). Hence, the bandwidth of the system will get increased.[3] A simple CM consists of 2 MOS transistors with compensating resistor R as shown in fig. 1.

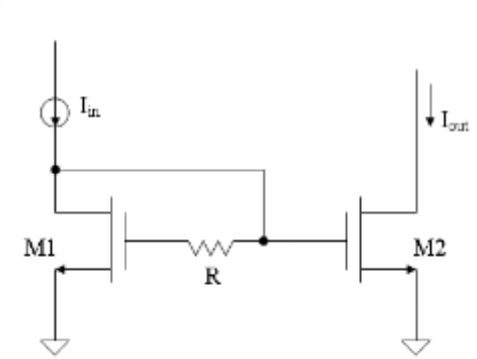


Fig. 1: Simple CM [3]

The transfer function (TF) of the current mirror without compensating resistor R[3] is:

$$A_i(s) = \frac{g_{m2}}{g_{m1}} \frac{1}{\left(1 + s\left(2C_{gs}/g_{m1}\right)\right)}$$ and bandwidth is:

$$\omega_0 = \left(\frac{g_{m1}}{2C_{gs}}\right)$$

where g_m is the transconductance and C_{gs} is the gate-source capacitance of each MOS transistor. With compensating resistor, TF of the current mirror is transformed:

$$A_i(s) = \frac{g_{m2}}{C_{gs2}} \frac{\left(s + 1/RC_{gs1}\right)}{\left(s^2 + \left(\frac{C_{gs1} + C_{gs2}}{RC_{gs1}C_{gs2}}\right)s + \left(\frac{g_{m1}}{RC_{gs1}C_{gs2}}\right)\right)} \quad (1)$$

The transfer function of the simple current mirror is transformed from first order single pole to second order low pass consisting of 1 zero and 2 poles. The zero and poles of the resistively compensated simple current mirror are:

$$Z_1 = -\frac{1}{RC_{gs1}} \quad \text{and}$$

$$P_{1,2} = \frac{C_{gs1}+C_{gs2}}{2RC_{gs1}C_{gs2}}\left[-1 \pm \sqrt{1-\frac{4g_{m1}RC_{gs1}C_{gs2}}{\left(C_{gs1}+C_{gs2}\right)^2}}\right] \quad (2)$$

From equation (1), the bandwidth of the system is obtained as (3):

$$\omega_0 = \sqrt{\frac{g_{m1}}{RC_{gs1}C_{gs2}}} \quad (3)$$

For $R = 1/g_{m1}$ and $C_{gs1} = C_{gs2}$, the zero gets cancelled with one of the poles resulting into a first order transfer function. The -3 dB frequency of the compensated CM[3] is: $\omega_0 = \left(\frac{g_{m1}}{C_{gs}}\right)$. It can be concluded that the -3 dB frequency of the resistively compensated CM is twice that of the previous one. Same method has been further applied for bandwidth increment of a FVF based cascode current mirror.

The conventional LVCCM

The conventional simple current mirror structure has drawbacks of low ratio of output to input impedance. Some circuits were reported earlier such as regular cascode current mirrors to improve the output impedance but suffer from increased minimum supply voltage which limited the applicability of these structures for low voltage operation.[11-14] In order to meet the present electronics industry requirements of low power supply, many circuits are available and the most commonly used cell is the FVF based LVCCM.[1,15-19] It can is seen that the performance of FVF based cascode current-mirrors (shown in fig. 2) including maximum operating signal and error of current transfer, is better in comparison to conventional LVCCM.

Thus, FVF based CMs could be used in good-performance and low operating voltage analog systems. If all transistors of the current mirror are in the saturation region, shunt feedback causes impedance at input node to be low. Thus, the current flow amount through this input node will not affect its voltage. Hence designer can achieve high performance current mirror.[10] Another CM topology which is extensively used in analog application is alternatively-fed FVF based cascode current mirror[1] as depicted in fig. 3.

The minimum required supply voltage (V_{DD}) and minimum output voltages (V_{out}) are expressed:[1,20]

$$V_{DD,\min} = V_{TH(Mn11)} + V_{DS,sat(Mp1)} \quad (4)$$

$$V_{in,\min} = V_{TH(Mn11)} + V_{DS(Mn5)} \quad (5)$$

$$V_{out,min} = V_{DS,sat(Mn2)} + V_{DS,sat(Mn3)} \quad (6)$$

In order to further decrement of the input resistance (R_{in}), the topology of FVF based cascode CM is modified in a way that input current is fed at a different node i.e. output node of the input FVF [1]. The minimum required V_{DD} and minimum V_{out} are obtained as: [1,20]

$$V_{DD,min} = V_{TH(Mn1)} + V_{DS,sat(Mp1)} \text{ and } V_{in,\min} = V_{DS,sat(Mn11)}.$$

It can be noticed from equations (9) and (12) that there is a significant reduction in input voltage. The Rin of the FVF based cascode CM is given by the expression:

$$R_{in} \cong \frac{2}{g_{m1}g_{m5}r_{o5}}$$

where gmi (i=1, 5) is the transconductance of Mi and r_{o5} is the output impedance of the M5 transistor.

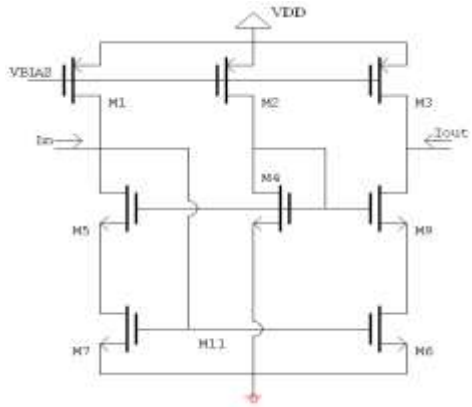


Fig. 2: FVF based LVCCM[1]

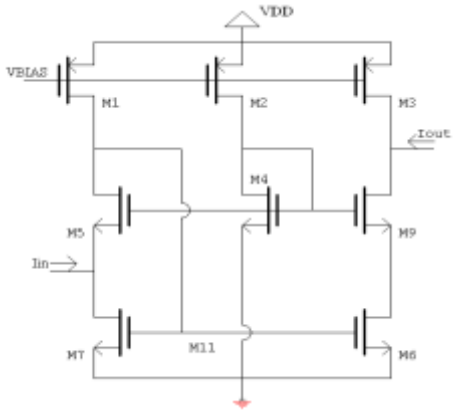


Fig. 3: Alternatively fed FVF based LVCCM

Resistively compensated FVF based LVCCM

The resistively compensated low voltage FVF based CM has been designed. The wideband flipped voltage follower has been inserted at the input terminal of the LVCCM shown in fig. 4 to enhance the bandwidth of the circuit.[10] The modified CM is shown in fig. 5 (R_{COMP} and R are compensating and feedback resistance respectively). Fig. 6 shows the actively compensated CM suggested by Gupta et al[10] and the modified version of fig. 6 is shown in fig. 7 (M_{COMP} is the transistor used for active compensation).

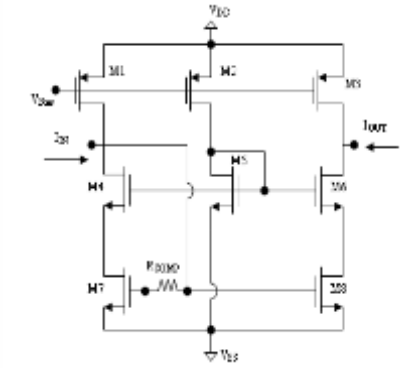


Fig. 4: Passively compensated CM[10]

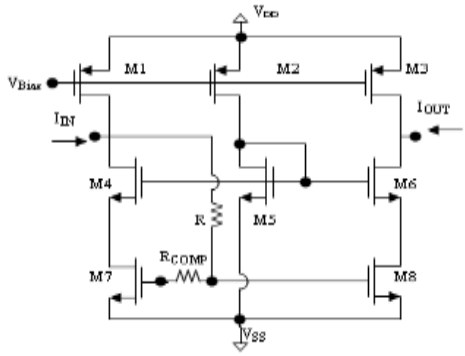


Fig. 5: Modified CM with the resistively compensated FVF

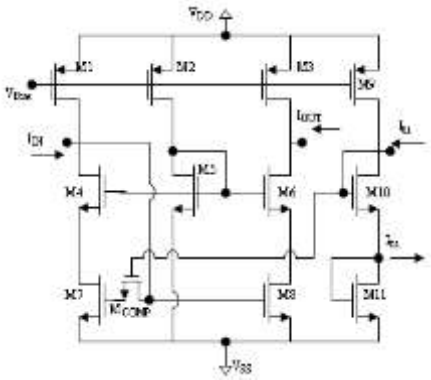


Fig. 6: Actively compensated CM[10]

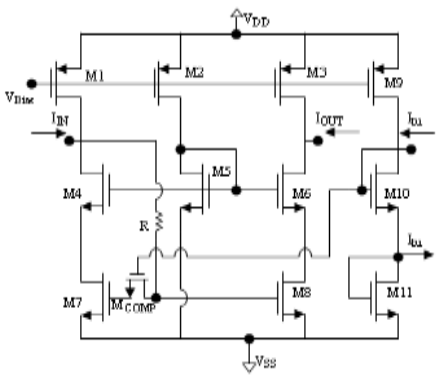


Fig. 7: Modified actively compensated CM with the resistively compensated FVF

Simulation Results and Discussion

Spectre simulator of Cadence using model parameters of TSMC 0.18 μm CMOS BSIM 3 and level 49 technology has been used to authenticate the functionality and performance development of all analog circuits. The simulation results of all the compensated LVCCMs are shown here.

The error (Iout-Iin) is shown in fig. 8 and it is almost -3.37%. The input and output compliances are portrayed in figures 9 and 10 respectively. From fig. 11, it can be seen that at 50 μA input current the modified CM dissipates 324.3 μW. It can be seen that when CM circuits are modified, the DC performance factors do not vary. Fig. 12 and 13 show the effect of resistive compensation on both the resistances of CMs. The obtained input and output resistances of CMs are 1.935 kΩ and 0.22 MΩ respectively. It can be observed from fig. 12 that the Rin of the CM decreases with frequency as the value of feedback resistance R increases (conventional CM is shown by solid line, CM with R = 1.6 kΩ and 6 kΩ are shown by dotted line with cross marker and dashed line with circle marker respectively).

Therefore, it leads to enhanced current flow at node output of FVF and input node of CM. From fig. 13, it can be seen that the output resistance of the designed LVCCM is same. Fig. 14 shows the frequency responses of CMs (fig. 2 and fig. 4 and 5). An improvement of 1.2 GHz in -3dB frequency is achieved by using the wideband FVF in passively compensated CM[10] i.e. BWER is 1.3 approximately.

However, peaking has been observed to achieve BWER of 1.6 which limits the maximum value of the compensating resistor. The frequency responses of the conventional and modified (fig. 6 and 7) actively compensated CMs can be depicted in fig. 14. The BWER is 1.2 of the improved actively compensated CM. It can be seen from figures 14 and 15 that the bandwidths of the passively compensated CMs are larger than that of the actively compensated CM. Active implementation of the compensating resistor provides several advantages such as smaller chip area requirement, but it provides smaller bandwidth than a passive resistor and increases peaking in the frequency response.

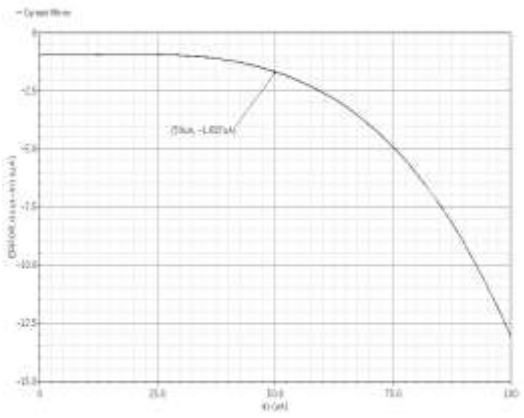


Fig. 8: Transfer error (Iout-Iin)

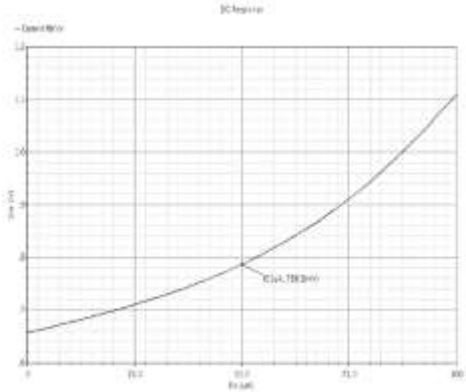


Fig. 9: Input Voltage (Vin) Versus Input Current (Iin)

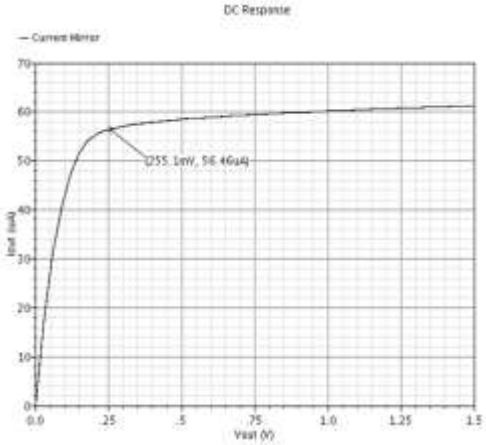


Fig. 10: Output Current (Iout) as a function of Vout

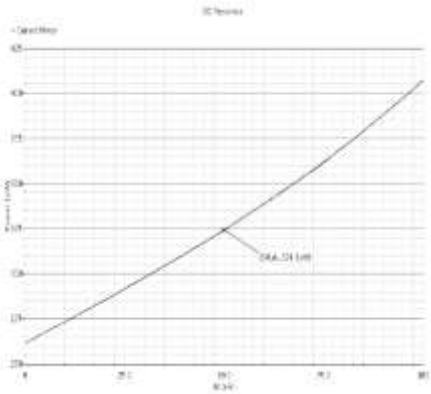


Fig. 11: DC power consumption

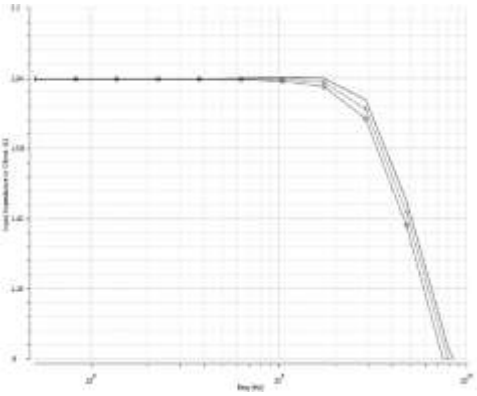


Fig. 12: Rin of (Fig. 2) and (Fig. 4 and 5) CMs

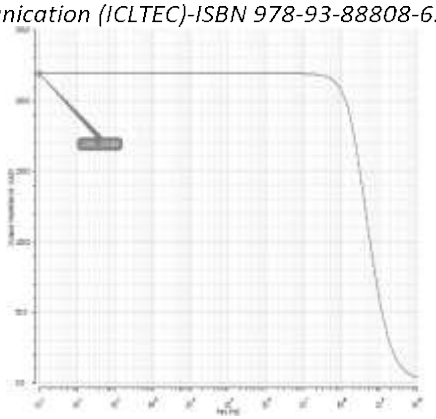


Fig. 13: Output impedance w.r.t frequency

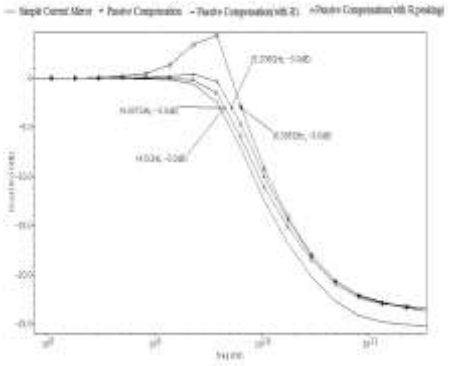


Fig. 14: Frequency responses of passively compensated CM (Fig. 5)

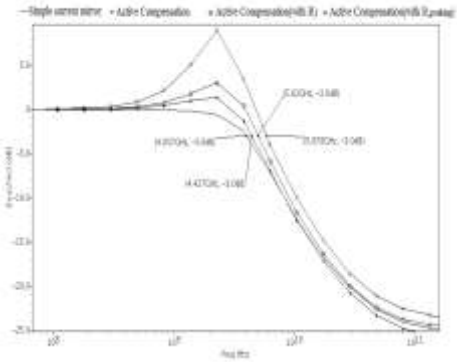


Fig. 15: Frequency responses of actively compensated CM (Fig. 7)

Conclusion

This work is dedicated to the development of FVF based conventional CM with lower power consumption. A resistively compensated FVF is used in place of conventional one in the low voltage CM to increase the -3dB frequency. The bandwidth of the proposed wideband passively compensated and actively compensated CMs is 6.395 GHz and 5.02 GHz respectively. From simulation results it has been inferred that the designed CMs circuits exhibit large bandwidth without any variation in the DC performances and hence, these circuits may find wide range of applications in high speed signal processing systems.

References

1. Koliopoulos C. and Psychalinos C., A Comparative study of the performance of the flipped voltage follower based low voltage current mirror, IEEE Int. Symp. Circuits and Systems, 1-4 **(2007)**

2. Carvajal R., Ramirez-Angulo J., Lopez Martin A., Torralba A., Galan J., Carlosena A. and Chavero F., The Flipped voltage follower: a useful cell for low voltage low power circuit design, *IEEE Trans. on Circuits and Systems*, **52(7)**, 1276 – 1279 **(2005)**

3. Voo T. and Toumazou C., High-speed current mirror resistive compensation technique, *IEE Electronics Letters*, **31(4)**, 248 – 250 **(1995)**

4. Voo T. and Toumazou C., Precision temperature stabilised tunable CMOS current-mirror for filter applications, *IEE Electronics Letters*, **32(2)**, 105-106 **(1996)**

5. Voo T. and Toumazou C., Efficient tunable continuous-time integrated current-mode filter designs, IEEE International Symposium on Circuits and Systems, 93-96 **(1996)**

6. Rajput S.S. and Jamuar S.S., A high performance current mirror for low voltage design. IEEE Circuits and Systems, In proc IEEE APCCAS, Tianjin, China 170-173 **(2000)**

7. Rajput S.S. and Jamuar S.S., Advanced current mirrors for low voltage analog designs, Proc. of IEEE ICSE '04, Kuala Lumpur, Malaysia, 258-263 **(2004)**

8. Rajput S.S. and Jamuar S.S., Low voltage, low power, high performance current mirror for portable analogue and mixed mode applications. Circuits, Devices and Systems, IEE Proc., 273 – 278 **(2001)**

9. Jamuar S.S. and Rajput S.S., Low-voltage current-mode analog circuit structures and their applications, IEEE International ASIC/SOC Conf., 477- 478 **(2002)**

10. Gupta M., Aggarwal P., Singh P. and Jindal N.K., Low voltage current mirrors with enhanced bandwidth, *Analog Integrated Circuits Signal Process*, **59**, 97-103 **(2009)**

11. Prodanov V. and Green M., CMOS current mirrors with reduced input and output voltage requirements, *IEE Electronic Letters*, **32(2)**, 104-105 **(1996)**

12. Ledesma F., Garcia R. and Ramirez-Angulo J., Comparison of new and conventional low voltage current mirrors, Midwest Symp. on Circuits and Systems, 49-52 **(2002)**

13. Yonghua C. and Geiger R.L., Cascode current mirrors with low input, output and supply voltage requirements, IEEE Midwest Symp. on Circuits and Systems, 490 – 493 **(2000)**

14. Palumbo G., On the high-frequency response of CMOS cascode current mirror, IEEE proc., 7th Mediterranean Electrotechnical Conf., 1211-1214 **(1994)**

15. Ramirez-Angulo J., Gupta S., Padilla I., Carvajal R.G., Torralba A., Jimenez M. and Munoz F., Comparison of conventional and new flipped voltage structures with increased input/output signal swing and current sourcing/sinking capabilities, *Midwest Symposium on Circuits and Systems*, 1151-1154 **(2005)**

16. Ramirez-Angulo J., Carvajal R.G., Torralba A. and Galan J., Low supply voltage high-performance CMOS current mirror with low input and output voltage requirements, *IEEE Trans. on Circuits and Systems-II*, **51(3)**, 124-129 **(2004)**

17. Meaamar A., Low-Voltage, High-Performance Current Mirror Circuit Techniques, IEEE International Conf. on Semiconductor Devices, Kuala Lumpur, Malaysia, 661-665 **(2006)**

18. Torralba A., Carvajal R.G., Ramirez-Angulo J. and Munoz E., Output stage for low supply voltage, high-performance CMOS current mirrors, *Electronics Letters*, **38(24)**, 1528 – 1529 **(2002)**

19. Aggarwal B., Gupta M. and Gupta A.K., A comparative study of various current mirror configurations: Topologies and characteristics, *Microelectronics Journal*, **53**, 134-155 **(2016)**

20. Razavi B., Design of Analog CMOS Integrated Circuits, McGraw-Hill, Inc. **(2001)**.

The Gas Leak Detection Based on a Wireless Monitoring System

K.SUHASINI, MTech scholor, Department of Electronics and Communication Engineering,

Hyderabad, India. suhasinikangati@gmail.com

Abstract—**The industrial gas leaks cause accidents and pose threats to the environment and human life. Thus it is essential to detect the gas leaks in time. Usually, the abnormal concentration signals are defined by a fixed concentration value, such as 25% of the lower explosive limit (LEL). However, it is difficult to accumulate to the fixed point quickly when the leak is small. In addition, the actual leak signals are seldom available, making many data classification inoperable. To solve these problems, this paper proposes a detection approach by using the auto-correlation function (ACF) of the normal concentration segment. The feature of each normal segment is obtained by calculating the correlation coefficients between ACFs. According to the features of statistical analysis, a non-concentration threshold is determined to detect the real time signals. In addition, the weighted fusion algorithm based on the distance between the sensors and virtual leak source (VLS) is used to fuse multi-sensory data. The proposed method has been implemented in a field by building a wireless sensor network (WSN). It is confirmed that the system detection rate reaches as high as 96.7% and the average detection time delay is less than 30s on the premise of low false alarm rate.**

Index Terms—**Gas Leak Detection, the Auto-correlation Function (ACF), the Correlation Coefficient, the Weighted Fusion, Wireless Sensor Network (WSN).**

I. INTRODUCTION

Globally, the gas, petroleum, chemical, metallurgy and other industries produce a large number of flammable and toxic gases as well as benzene and other organic vapors every year. Due to improper man made operation or equipment aging, a large number of gas leaks have occurred. Some leaks that are not detected or repaired in time have caused great economic losses, environmental pollution and even huge casualties. Statistically, 778 safety accidents occurred in China from January 2016 to January 2017. The explosions and poisonings resulted from the gas leakage accounted for 6.17% and 5.4% reference to the total accidents, respectively [1]. The usages of hazardous materials are strictly restricted for the safety by the State Administration of Work Safety. However, the usage of chemical and hazardous materials increases year by year. At present, a large number of commercial companies and research institutions engage in researching the detection of dangerous gas leak. In the market, the handheld devices are most commonly used. These devices have high accuracy and accurate positioning. However, the users need to go to the detected areas where the leaks are prone to occur while their safety and real time detection can't be ensured [2]. The mobile robot with vision and gas sensors can replace the workers to detect gas leaks and send the data to remote control center that is a novel detection method and protects the lives of workers [3]. However, this detection system is limited by the mobility of the robot and is not suitable for the monitoring area with uneven terrain and wide range. In addition, the wire monitoring system that is composed of high resolution sensors and control centers solves the problems of the security and real time issues by using the cables to transfer data and compromises on the inflexible installation and expensive maintenance [4].

Recently, the sensor technology, wireless communication and the embedded technology are developing rapidly. The wireless sensor network (WSN) with low cost, flexible installation, real-time continuous detection and other advantages is used widely in the monitoring of environmental parameter, structural health [5] and gas leak. Somov et al. [6] deployed a ZigBee WSN to monitor the gas leaks in the boiler facility. The system consists of 9 battery powered wireless sensor nodes and 1 network ZigBee coordinator while the system also has access to the Ethernet and GSM network to send messages. Generally, the system of detecting gas leak by gas sensors sets a few fixed leak concentration to confirm gas leak or not. Jelicic et al. presented a wireless sensor network (WSN) for monitoring indoor air quality. The sensor node is designed with very low sleep current consumption and the network is multimodal. It exploits information from auxiliary sensors and neighbor nodes about gas concentration to modify the behavior of the node and the measuring frequency of the gas concentration.[7] Rossi et al. exploited the transient response of the sensing element, so that indoor air quality assessment power required is

This work is supported in part by key research and development plan project of Zhejiang province (No.2018C01036), the National Natural Science Foundation of China (No. 61871167 and 61376117). (Corresponding author: Linxi Dong and Gaofeng Wang)

Linxi Dong is with the School of Electronics and Information, Hangzhou Dianzi University, Hangzhou 310018, China, the State Key Lab of Functional Materials for Informatics, Shanghai Institute of Microsystems and Information Technology, CAS, Shanghai 200050, China, and also with the State Key Lab for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China (Email: donglinxi@hdu.edu.cn). Libo Zhao is with State Key Laboratory for Manufacturing Systems Engineering, School of Mechanical Engineering in Xi'an Jiaotong University , China(Email: libozhao@xjtu.edu.cn). Zhiyuan Qiao, Haonan Wang, Weihuang Yang, Wensheng Zhao, Kuiwen Xu, Gaofeng Wang are all with the Key Laboratory of RF Circuits and System of Ministry of Education, Electronic and Information College of Hangzhou Dianzi University, Hangzhou 310018, China. (Email: 2569173072@qq.com,1142214134@qq.com,89297508@qq.com,wshzhao@hdu.edu.cn,kuiwenxu@hdu.edu.cn,gaofeng@hdu.edu.cn). Haixia Yan is with School of Information Engineering, Hangzhou Dianzi University, Hangzhou 310018, China (Email: 1328515466@qq.com)

reduced to 1/20 of the original.[8] In the paper [9], two alert levels that are high alert and low alert are set respectively. According to the setting specifications for safety monitoring and controlling equipment in dangerous chemicals major hazard installations, two fixed points are set at 25% of the lower explosive limit (LEL) and 50% LEL usually. However, the leaks cannot be detected by setting a fixed concentration threshold generally when the leakage is small or just starts to happen.

Different from the common detection method of gas concentration for the leaks, the methods of infrared and ultrasonic detection develop rapidly in recent years. The former method uses thermal imaging and infrared image processing to detect leaks [10]. For example, a low cost infrared camera called SENSIA's Gas Imaging System [11] developed by Universidad Carlos III de Madrid is a low cost infrared camera based on spectrally adapted and high sensitivity imaging technology. It can easily detect and identify fugitive gas emissions. Additionally, Leis *et al.* proposed improvements in the detection time [12] and the optical flux variation due to using solid-state IR sources to heat [13] respectively, which promoted the development of infrared sensors in leak detection. The ultrasonic method detects the leaks by analyzing the ultrasound that generated under pressure. The approach is not affected by inclement weather, wind direction, and can respond quickly without physical contacts. The GDU-Incus ultrasonic gas leak detector released by the United States Emerson Group can detect the gas leak in the range from 2 to 40 meters [14]. Also, it is unnecessary to wait until the gas concentration accumulates to the dangerous threshold. In the actual environment, the two kinds of devices are often set on the towers in order to monitor a larger area. Both of them are difficult to detect the leaks with low leakage pressure, the environmental obstacles, or interferences of heat sources.

In order to solve the small leak detection, many algorithms have used in the field of pipeline leak detection. The probabilistic and data classification methods are the common approaches. Akouemo *et al.* [15] used a linear regression model and a geometric probability distribution of the residuals to determine the anomalous probability of a data, and then trained a Bayesian maximum likelihood classifier to distinguish between false positives and true anomalies. Gupta *et al.* [16] proposed a probabilistic method which used Bayesian probabilistic framework and the steady state flow and pressure values of gas to detect the presence of a leak event. In addition, Wang *et al.* [17] proposed a pipeline leak detection approach by using time-domain statistical features from normal sample signals, and built the support vector data description (SVDD) model to finish the leak detection. For the small leaks, some results have been achieved. Xiao *et al.* [18] used the variational mode decomposition (VMD) to do the components reconstruction and proposed the ambiguity correlation classification (ACC) based on the correlation coefficient to detect the small leak of pipeline. A novel de-noise algorithm based on dual tree complex wavelet transform and singular value decomposition (DTCWT-SVD) is applied for small leak detection [19]. Similarly, the harmonic wavelet based pipeline small leakage detection method is proposed by Hu *et al* [20]. Furthermore, Kang *et al.* [21] presented a water leakage monitoring architecture using the one dimensional convolutional neural network and a support vector machine (1D-CNN-SVM) that utilized the feature maps of CNN as input to the classifier and the approach does not need data transformation.

For the small leaks, it is difficult for using the concentration threshold distinguish the small leakage. So the features should to be extracted in the term of time-domain or frequency-domain so that the abnormal signals are distinguished. In the field of pathological analysis of electrocardiogram (ECG), an approach is proposed which extracted the Auto-correlation Functions (ACFs) of segments and analyzed the similarity between the ACFs of normal and abnormal to define the ECG signals [22]. The essence of this method is the thought of random signal processing [23]. However, the approach is to analyze whether the historical data contains anomalies after the sampling of signals have been completed. Therefore the approach cannot be directly used for the online leak detection. Pister *et al.* [24] proposed a method combined the thought in ECG and the likelihood function to detect the industrial gas leaks. By establishing the concentration distribution model, the likelihood function is obtained, and then the result of leak detection is got by the method mentioned in the ECG. However, the threshold is obtained through constantly trying, and the impact of sensor node's position was not considered.

To tackle these limitations, this work further advanced the previous method from the following perspectives:

- Implement an automatic procedure to select sensor nodes that participate in leak detection, considering the location of sensors;
- Explicitly define the degree of the auto-correlation function (ACF) as the feature for leak detection;
- Propose a general procedure for the threshold that distinguishes between abnormal and normal;

In the proposed method, normal data are sampled under the normal operation environment. The sample is first divided into fixed length segments. Then, the ACF of each segment is obtained and the degree of each segment is calculated as the time-domain feature by computing the correlation coefficient between each pair of ACF. It is employed as the discriminative feature for leak detection. Furthermore, the threshold, which is defined as the baseline to detect the abnormal signal, is obtained by analyzing the distribution of the degrees of the normal signals.

In this work, the data collected by sensor nodes firstly are initially analyzed and classified. The concentration is divided into three levels: the first level is greater than 50% LEL; the second level is within 25-50% LEL; the third level is less than 25%LEL. The first two cases are defined as concentration abnormal while the third level is undefined. In addition, the temperature, pressure and humidity are divided into normal and abnormal according to fixed alarm values, respectively. The system in this study will sound alarms immediately when the abnormal results are obtained. And the undefined concentration data will be defined by the method proposed in this paper. The reliability is verifed by comparing the measurement data of the professional sensor equipment with the data in our system.

The rest of the paper is organized as follows. The wireless monitoring system for the gas leak detection is introduced in Section II. Section III describes the method of the gas leak detection including the selection of the multisensory data and the feature extraction in time-domain mainly. The experimental

results are presented in Section IV. And this paper is summarized and concluded in Section V.

II.WSN Based Detection System

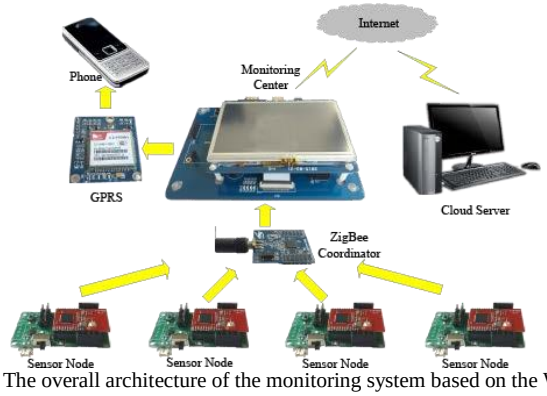


Fig. 1. The overall architecture of the monitoring system based on the WSN for the detection of gas leak.

Fig. 1 illustrates the overall architecture of the monitoring system based on the WSN for the detection of gas leak in an open environment. In order to detect the gas leak, a number of ZigBee nodes with gas sensors and a ZigBee coordinator need to be deployed to form a WSN in the monitored area. After converting and filtering, the data of gas concentration will be sent to the ZigBee coordinator. Then, the sensor nodes which send the third level data are selected to form a set by the ZigBee coordinator. The ZigBee network uses the 2.4GHz band. And the data of the sensor nodes in the set is fused into one concentration sequence (CS). Then the CS is transferred to the monitoring center of system and judged whether the gas leaks by the detection algorithm. The method used by this work is described in chapter III. Once the concentration signals are detected as abnormal by the detection algorithm, the GPRS module will send out an audible alarm and warning messages to the maintenance person. It works in the 900MHz band. It will also send the collected data and the diagnostic result to the cloud server via Internet.

Wireless Sensor Nodes Design

All the data analyzed by the detection system is acquired by the sensor nodes. So the impact of the environment factors on the sensor data will be mapped to the result of the detection system. At this point, the temperature, humidity and pressure sensors (THPs) are included in the nodes. The sensor nodes in this work are designed by MCU, ZigBee wireless transceiver module, sensors module, and power supply module.

Considering the feasibility and safety of experimental operations, ethanol is used as the test gas and the sensor TGS2620 was selected, with a 50-5,000 part per million (ppm) measurement range and a 0.3-0.5 sensitivity (resistance ratio). The sensor is comprised of a metal oxide semiconductor layer formed on an alumina substrate together with an integrated heater. In the presence of a detectable gas, the sensor's conductivity increases as the gas concentration increases. A simple electrical circuit can convert the change in conductivity to an output signal which corresponds to the gas concentration. For the test, it consumes an average of 42mA current at 5V. The power consumption of gas sensor is approximately 210mW. To meet power requirement of sensor node, it would be possible to utilize a solar scavenger with a rechargeable battery, and combined with adjustable sampling frequency according to gas leak or not, and unnecessary alarms report to minimize the energy consumption. As efficient microcontrollers become cheaper and less power hungry, the only component left to be improved is the sensor [24]. The main weakness of metal oxide semiconductor gas sensor used in the paper is its large energy consumption, which is also common shortcoming of current gas sensors. Low-power gas sensors are currently less commercially available, but the gas sensors reported in the literature have been able to achieve very low power consumption. For example, Yeon Hoo Kim et al. [25] demonstrated self-activated transparent all-graphene gas sensor, the power consumption is 12μW to 14.2mW for the applied voltage of 1 to 60V. The power consumption of Low power wireless gas sensors used by Andrey Som et al. [26] is 75mW in the continuous measurement. Its low power consumption is achieved by applying a heater implemented as 10 μm platinum microwire in glass insulation. The literature [27] realized gas sensor by integrating the novel 3D hybrid aerogel on a low-power Micro-heater platform, and when the temperature of Micro-heater is 200^{o}C, the power consumption is only 4mW. In the future, the commercial application of these low-power gas sensors will greatly improve the continuous working time of wireless gas sensor nodes.

The scholars also researched the other strategies to solve the energy consumption and power supply problems of wireless sensor nodes. The literature [28] studied the system- and circuit-Level optimization of power supply system for Wireless Sensor Networks, and renewable power-supply system. Mingyi Chen [29] presented a self-powered wireless sensor node powered by electromagnetic energy harvester, the results show that designing chip architecture with less components is also good strategy.

The algorithm mentioned in the paper is suitable for low sampling rates to conserve the AD sampling's power consumption and when detecting a sudden gas leak, then increase sampling frequency to ensure the sensor data accuracy. To increase the continuous working time of wireless gas sensor node, choosing the large capacity lithium cells is also one compromise method. For example, when the gas sensors work intermittently, and it works 10s every 20 seconds for data acquisition when have no gas leakage. Based on the 10000mAh size lithium cells available on the market, the gas sensor can work for about 40 days. If the photovoltaic solar panel and future low power consumption gas sensors with a few μw to mw order of power consumption will be used in the sensor nodes, a longer working time can be reached.

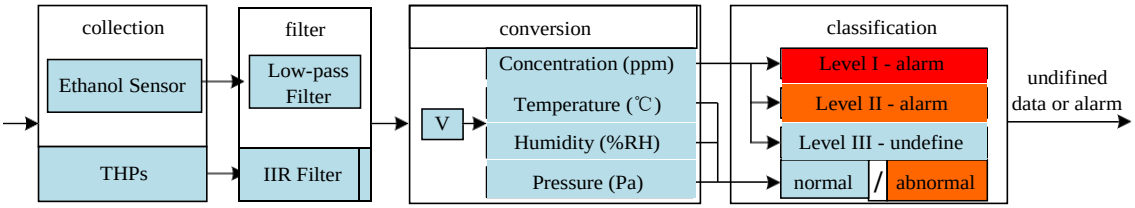


Fig. 2. The flow diagram of the data processing in the sensor nodes.

For the THPs, the BME280, which is a MEMS sensor integrated temperature, humidity and pressure, is adopted. The sensor has a built-in IIR filter to filter short-term disturbances. It consumes 11.88μW when it works and 0.33μW when it is dormant. For the MCU and wireless transceiver module, the highly integrated CC2530F256 chip is used as the node processor and transceiver. In order to achieve the minimum power consumption of the node, we reduce its transmission power under the condition of satisfying the normal monitoring of the system, where the transmission distance is 70 m. The peak of working current specified for CC2530F256 chip is 79 mA at 3.3 V supply in continuous measuring mode. The major function of this chip includes the collection of signals, filtering, Voltage/Indicators conversion, classification and the signal transmission. The flow diagram of the data processing is shown in Fig. 2.

In addition, the sensor node is powered by a lithium battery and the BP24210 solar charging module produced by TI. A photography of the sensor node can be seen in Fig. 3.

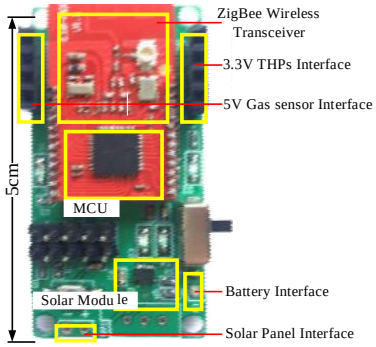


Fig. 3. The physical diagram of the sensor node.

The Deployment of Sensor Nodes

The placement of the sensor is critical to the detection result. It affects the collection of data directly, even if the sensor is located near the gas source. The sensor nodes can be distributed randomly, deterministically, or uniformly. On the premise of being familiar with the monitoring area, the deployment of sensors is deterministic. The sensor nodes should be deployed at the location where the concentration data is prone to be collected. The simulations for the leak of volatile ethanol by Fluent are performed to analyze the distribution of the concentration and further determine the position of the sensor under different conditions. Fig. 4(a) presents that the leaking gas is influenced by the wind direction, the source nozzle size is 6mm and the wind direction is set from left to right. It is observed that the gas diffusion is divided into four stages: first, the gas diffuses in the initial injection direction; second, the gas spreads mainly in the downwind direction under the influence of the wind; then the gas sinks and accumulates as the density

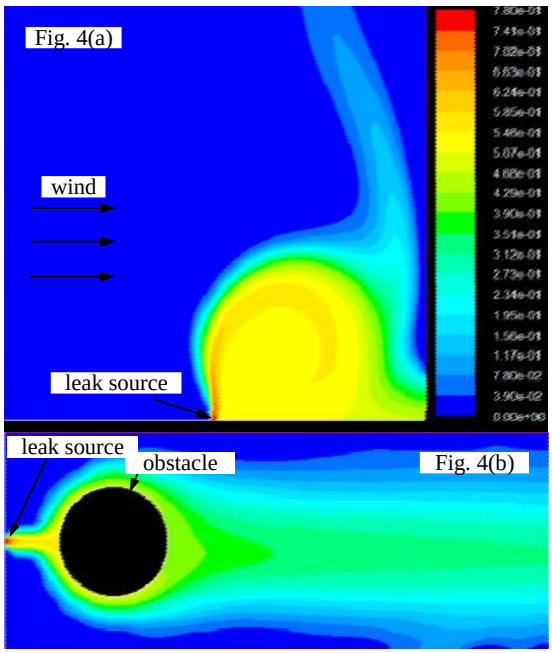


Fig. 4(a). The gas diffusion simulation under the influence of the wind direction and gas characteristics.
Fig. 4(b). The concentration distribution in presence of an obstacle near the source of the leak.

of ethanol is greater than the air's; third, the buoyancy force plays a leading role and the gas diffuse upward; finally, the gas concentration is negligible when the gas is diluted in air. And Fig. 4(b) illustrates the concentration distribution in presence of an obstacle near the leak source. The leaking gas bypasses the obstacle or accumulates nearby the obstacle. There has higher concentrations at both front and back of the obstacle.

From Fig. 4(a), the sensor is mainly collected during the second stage of gas diffusion, since the initial injection direction is unpredictable and the concentration is too low in the latter two stages.

Therefore, combined with the simulation results and the safety specification of combustible gas detection, the deployment rules of the sensor nodes are summarized as follows:

- The more sensor nodes should be installed at the downstream of the high frequency wind direction relative to the potential leak source;
- When the density of the gas is greater than 0.97kg/m^3, the sensor nodes should be deployed below the level of the potential leak source;
- Considering the presence of the obstacle, the sensor should be placed in the vicinity of the obstacle (e.g. columns or walls);
- A suitable obstacle will be placed near the sensor nodes to increase the gas residence time when the monitored area is open or the wind speed is high;

III. The Method of the Gas Leak Detection

A. Time-Domain Features Extraction

Time-domain features such as variance, mean, kurtosis or skewness are not suitable as a basis for judging between normal and abnormal. The main drawback of mean is that it is easy to be affected by extreme values. By increasing the time window of each average, the effect of extreme values can be reduced, but the monitoring delay of the system is increased. Variance can indicate the deviation of all samples from the mean value to a certain extent, but it can not be used as a parameter for rapid detection of gas leakage. Kurtosis are statistics describing the overall extent of the slow distribution of all forms steep values, which is not suitable as a statistical parameter for gas leakage. Skewness is a measure of the skewness and extent of statistical data distribution. It is a numerical feature of the degree of asymmetry of statistical data distribution, and is not suitable for dynamic feature statistics. Mean and variance is used in ACFs. They are part of the algorithm. These parameters, such as mean, variance, is not suitable for determining whether a gas is leaking or not. The algorithm is supposed to detect the leaking as soon as possible, but these parameters are time-consuming.

When no leaks occur, the concentration of the monitored gas approximates to zero. And if leaks happen, the amplitude of concentration will fluctuate and deviate from zero. Therefore, a common detection method of gas leak is to set fixed alarm concentration point (such as 25%LEL and 50%LEL). The problem is that the concentration is difficult to accumulate to the alarm points when the leak aperture is small. Thus, the non-concentration threshold need to be set.

The measurement curves of gas sensors in the process of gas leakage and non-leakage， as well as the sampled normal and abnormal signal of concentration in three minutes are shown in Fig. 5. According to product information of TGS 2620, this sensor is insensitive to low level ppm of ethanol. When the gas concentration is lower than 50ppm, the sensor has a lower precision because of the insensitivity. The sampling rate of the sensor is 1Hz during the experiment. There is no need for high sampling rate in the detection process of the system, which is helpful to reduce the power consumption. The abnormal signal was generated with a man-made ethanol gas leak, whose aperture is about 6mm. The ethanol sensor is placed downwind at a distance of 20cm from the leak source made by our laboratory.

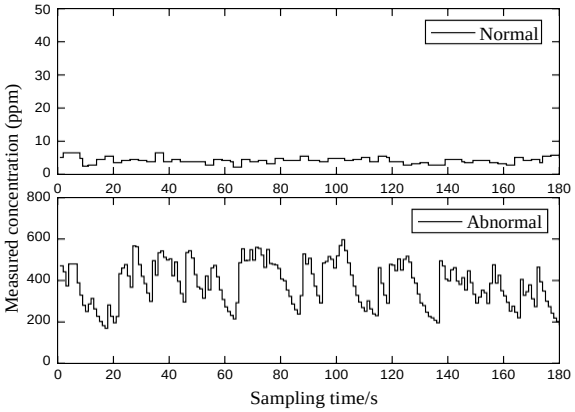


Fig. 5 Top: normal concentration data (without leak); bottom: abnormal concentration data (with leak).

The signals of normal and abnormal are divided into three segments respectively, and then the ACF of each segment is calculated. The result is shown in Fig. 6.

The similarity between ACFs is determined by the correlation coefficient, which is quantified by:

$$rr_{iiii} = \frac{CCCCCC(AA_{ii}, AA_{ii})}{\sigma\sigma_{ii} \cdot \sigma\sigma_{ii}} = \frac{EE((AA_{ii} - \mu\mu_{ii})(AA_{ii} - \mu\mu_{ii}))}{\sigma\sigma_{ii} \cdot \sigma\sigma_{ii}} \quad (1)$$

where AA_{ii} and AA_{ii} are two ACF, $\mu\mu_{ii}$ and $\mu\mu_{ii}$ are the means of AA_{ii} and AA_{ii}, $\sigma\sigma_{ii}$ and $\sigma\sigma_{ii}$ are the standard deviations of AA_{ii} and AA_{ii}. In this way, the correlation coefficients rr_{iiii} are normalized to the range of 0 to 1 firstly and then a correlation matrix composed of rr_{iiii} is formed. Each segment is characterized by a degree value, which is obtained by calculating the mean of the coefficients for each column of the matrix. When the segments are all normal, the degrees of them are high. And when the segments contain many normal segments and an abnormal segment, the degree of the abnormal segment is low. The degree feature is independent of the absolute amplitude and reflects the similarity between an undefined segment and normal ones. Therefore,

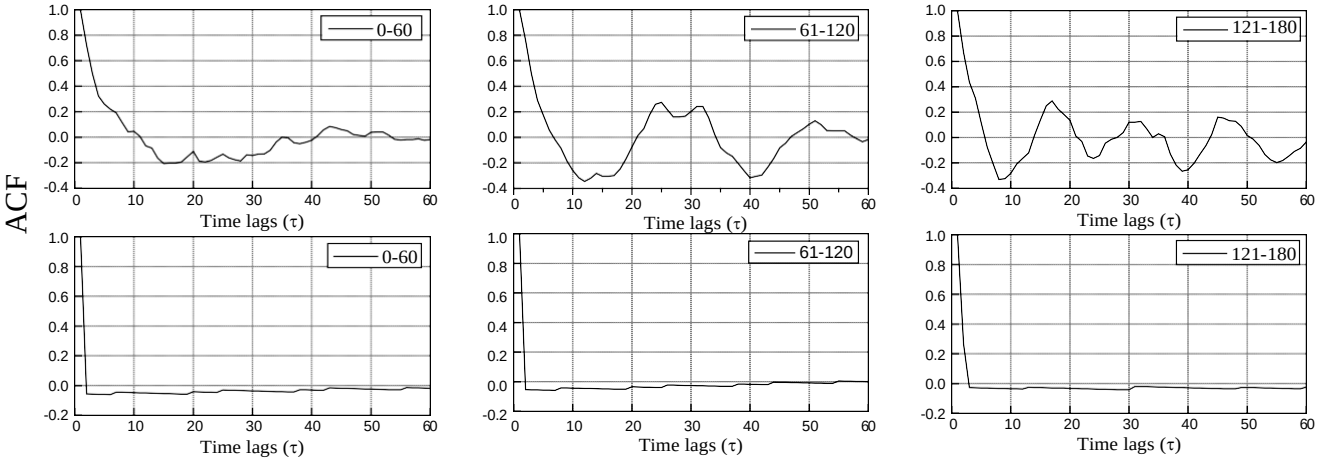


Fig. 6. ACFs of six different concentration segments of on minute. When the signal is normal, their ACFs are similar and stable while the abnormal ACFs have large fluctuations which are different strongly from the normal ACFs. This can be seen in Fig. 10 where the ACFs of three normal segments (bottom) and three abnormal segments (top) are plotted together.

the degree can be used as the feature to differentiate the normal and abnormal signals.

B. Implementing Leak Detection with Time-Domain Features

1) The General Procedure: For the system implementation, the leak detection model is established with the normal concentration data and applied to the real-time gas leak detection. Fig. 7 illustrates the leak detection procedure used by this study.

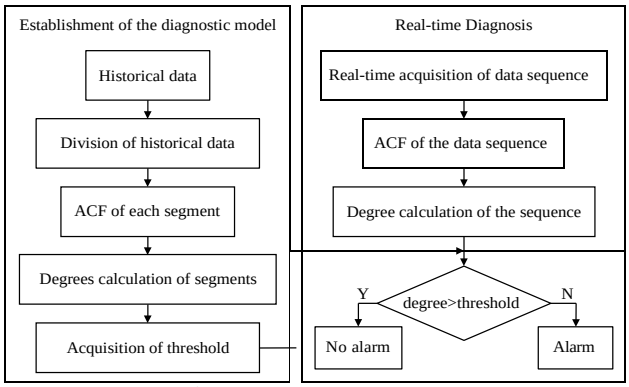


Fig. 7. The procedure for real-time leak detection

2) Feature Extraction: The training historical data took two hours to complete the sampling continuously when no leak. The time-domain features, which described in section III-A, can be extracted by the following steps.

a. Divide the historical data into *M* segments with a step value *N* (with *N*=30).
b. Then the ACF of each segments is calculated.
c. Next, the correlation matrix is formed according to the equation (1).
d. Finally, the degrees RR_{ii} of ACFs are computed by averaging each column of the correlation matrix:

$$RR_{ii} = \frac{1}{MM-1}\sum_{ii=1}^{MM-1} (ii \neq jj) \qquad (2)$$

3) Threshold Determination: In the previous step, nearly three hundred degrees have been obtained. These degrees are used to do the demonstration of the method and more historical data is needed in practice. The histogram of the degrees' distribution can obtained by the following steps:

a. Divide the value of degrees into MM_0 groups with a step value $AA_{sssssss}$, which can be written as:

$$MM_0 = \frac{KK_{mmmmmm} - KK_{mmiimm}}{AA_{sssssss}} \qquad (3)$$

Where RR_{mmmmmm} and RR_{mmiimm} are the maximum and minimum of the degrees, respectively. The value changes in the second place after the decimal point, so the step $AA_{sssssss}$ is set to 0.05;

b. Count the number of degrees whose value fall in the range of the kkth ($k = 1, \cdots, MM_0$) group and scale it as frequency;
c. Calculate the fitting function of the histogram distribution based on the statistics of the degrees;

Plot the statistical distribution curve of the degrees and the fitting function;

Through the process of the fitting, the probability density function (PDF) of the degrees is obtained as is shown in Fig. 8.

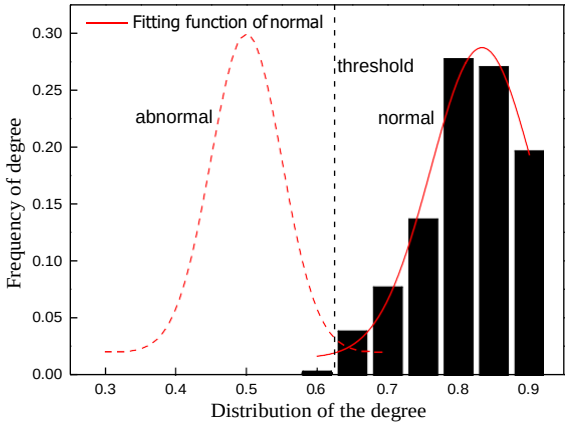


Fig. 8. The statistical distribution curve of degrees.

In other words, the probability $PP(RR_{ii}|\theta\theta = 0)$ against the degree RR_{ii} is obtained when no leak ($\theta\theta = 0$). According the previous theories described in section III-A, the degrees of the abnormal segment are less than the degrees of normal, generally, as is shown in the Fig. 9.

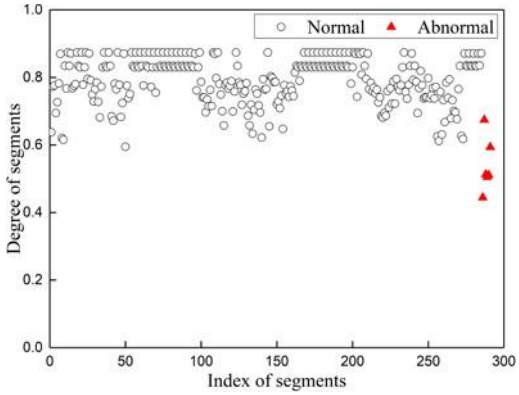


Fig. 9. The circle symbols represent normal degrees, while the triangle symbols represent abnormal degrees.

If enough leak data is available, the probability $PP(RR_{ii}|\theta\theta = 1)$ against the degree RR_{ii} also can be obtained when leak which is shown by the dashed line in Fig. 8. The sequential probability ratio (SPR) can be used to detect the leak [30]. However, it is almost impossible to get actual leak samples. So the threshold should be set so that the probability $PP(RR_{ii}|\theta\theta = 0)$ is low under the premise of higher detection rate and lower false positive. When the degree is the threshold, the probability $PP(RR_{ii}|\theta\theta = 0) = P_{sh}$.

The threshold is a critical parameter. The false negative judgment increases when the threshold is too low, while the false positive one rises when the threshold is too high. So a suitable balance between them needs to be confirmed.

4) Selection of Sensor Nodes and Fusion of the Real-time Data: Generally, the WSN uses a large number of sensors to obtain data. As the deployment grows in size, a faraway sensors from the leak source will probably not be able to detect any change. The response of the sensors to the plume of gas will depend on its location with respect to this plume, to confirm it, a test is completed by placing six sensors at different distances from the source of the leak in the downwind direction. It is found that the concentration decreases and phase of the detected concentration curve delay away from the source of the

leak, as shown in Fig. 10.

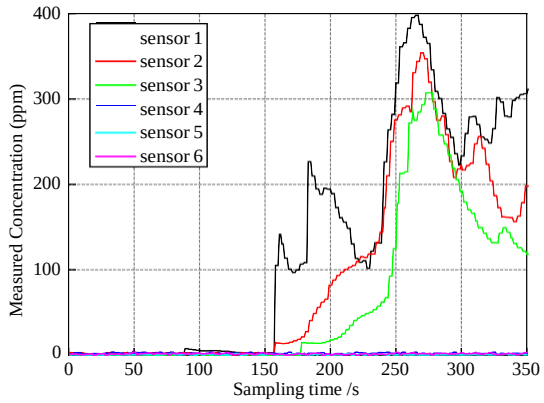


Fig. 10. The concentration curve of sensors at different distances from the leak source.

The several sensors closest to the leak source should make up a set to participate in the detection of gas leak such as sensor 1, 2 and 3. However, the location of the source leak is uncertain. The general rule is that the closer the distance to the leak source, the greater the gas concentration. So the sensor with the largest mean concentration over a period of time was assumed to be the virtual leak source (VLS).

In this work, the concentration data collected by the iith sensor at time tt is expressed as $xx_{ii}(tt)$ where i $\in \{1,2,\ldots,\mathrm{n}\}$, n is the total number of sensors that detected the third level data. The distance $dd_{iiiiiii}(tt)$ between the VLS sensor and sensor i can be written as:

$$dd_{iiiiiii}(tt) = \exp\left(-\left(xx_{ii}(tt) - xx_{iiiiii}(tt)\right)^2\right) \quad (4)$$

The smaller $dd_{iiiiiii}(tt)$ is, the greater the distance is. In addition, the distance $dd_{iiiiiii}(tt)$ is also fluctuant since the concentration is fluctuant. At a moment, a sensor is close to the VLS sensor, and it may be far away from the VLS sensor at the next moment. So when the mean of $dd_{iiiiiii}(tt)$ is large and its variance is small, the sensor i is considered close to the leak source in the period of time T. The distance $DD_{iiiiiii}$ is given as:

$$DD_{iiiiiii} = \lfloor 1 - 0.05\sigma\sigma_{ii}^2 \rfloor \cdot \overline{dd}_{iiiiiii} \quad (5)$$

And there are

$$\overline{dd}_{iiiiiii} = \frac{1}{NN}\left(dd_{iiiiiii}(tt_1) + dd_{iiiiiii}(tt_2) + \cdots + dd_{iiiiiii}(tt_{NN})\right) \quad (6)$$

$$\sigma\sigma_{ii}^2 = \frac{1}{NN}\sum_{kk=1}^{NN}\left(dd_{iiiiiii} - dd_{iiiiiii}(tt_{kk})\right)^2 \quad kk \in (1, NN) \quad (7)$$

Where N is the step value and the number of sampling points in the period of time T.

Thus, a set of distance is given as:

$$DD_{mm} = \{DD_{1iiiiii}, DD_{2iiiiii}, \cdots, DD_{mmiiiiii}\} \quad (8)$$

According to the value of $DD_{iiiiiii}$, the top m sensors are selected into the new set DD_{mm} that is given as:

$$DD_{mm} = \{DD_{1iiiiii}, DD_{2iiiiii}, \cdots, DD_{mmiiiiii}\} \quad (9)$$

Correspondingly, the set SS_{mm} including s sensors is obtained as:

$$SS_{mm} = \{ss_1, ss_2, \cdots, ss_{mm}\} \quad (10)$$

The weighting coefficient cc_{ii} of each sensor in the set SS_{mm} is calculated according the $DD_{iiiiiii}$ in the set of DD_{mm} as:

$$cc_{ii} = \frac{DD_{iiiiiii}}{\sum_{ii=1}^{mm} DD_{iiiiiii}} \quad (11)$$

Hence the data from the sensors close to the VLS play a more important role in the leak judgment. Finally, the value $\mathrm{X}(tt)$ in the CS, the result of fusion, is given as:

$$\mathrm{X}(tt) = \sum_{ii=1}^{mm} cc_{ii}\, xx_{ii}(tt) \quad (12)$$

The CS is recognized as the real-time data sequence and is used as the input of the diagnostic model.

5) Leak detection: Once the real-time CS is obtained, its ACF is computed firstly. According to the flow in Fig. 11, the real-time degree needs to be obtained. At the same time, several ACFs of the normal sequence are required for the calculation of the real-time degree. The length of data in one detection, denoted as L, is divided into MM_1 segments (named window 1). As a result, the length of each segment is $NN = LL/MM_1$ (named window 2) which is the same as the value mentioned in the step 2. The first MM_1-1 ACFs are taken as the reference values and the M_1th is the real-time ACF. So its degree is obtained according to the equation (1) and (2). If the new degree is less than the threshold, it means the input concentration is abnormal; otherwise, the concentration is normal.

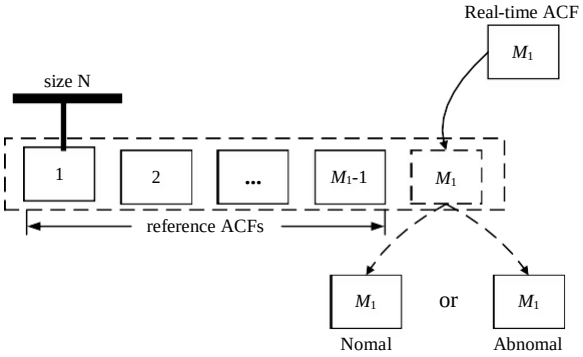


Fig. 11. The diagram of the real-time detection.

IV. Experimental Validation

A. *Verification of the effect of ambient temperature on the output response of the sensor*

As MOX sensor, the output of the gas sensor also depends on the temperature of the internal heater, we have done some experiments on the gas to analyze the effect of ambient temperature on the output response of the sensor. We have measured internal heater temperature and the response of the gas sensor at multiple temperatures in the temperature box. Experimental devices are shown in Fig. 12. Experimental device structure diagram is shown in Fig. 13.

The effect of working environment temperature on MOX gas sensor is shown in Fig. 14 with a red line. When the gas sensor is placed in the thermostat, the temperature of the gas sensor heater rises as the temperature of the thermostat rises, when the ambient temperature reaches 47.5℃, the heater temperature rises to 82.5℃. The output voltage of the sensor corresponding to the test temperature is shown in Fig. 14 with a blue line. The experimental results show that the output variance of the sensor is 0.126 in the working environment temperature range of 25℃ to 47.5℃.

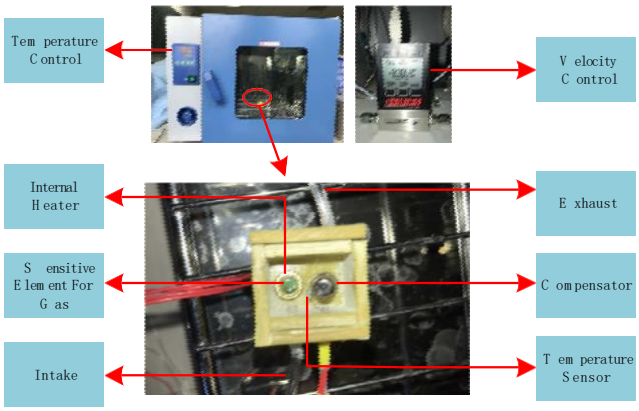


Fig. 12. Experimental devices

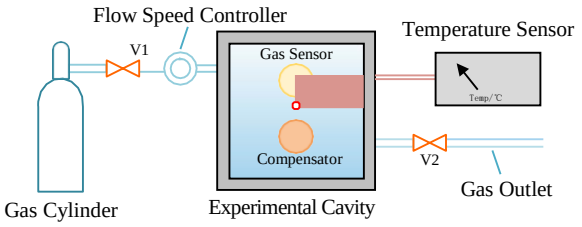


Fig.13 Experimental device structure diagram

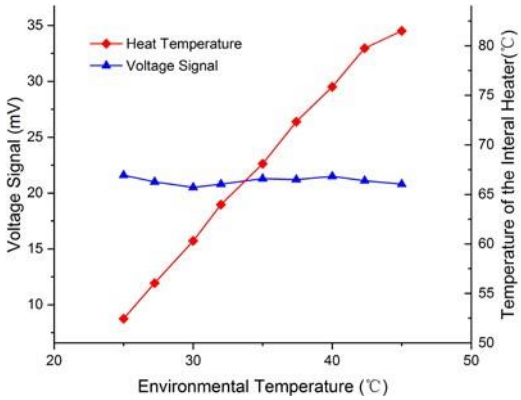


Fig. 14 Effect of ambient temperature on temperature of internal heater

It can be seen from the diagram that the heater temperature in gas sensor increases with the environment temperature, but the effect of environment temperature on the gas sensor is relatively small. This is because of the existence of compensator inside the sensor, the external temperature has little interference on the sensor. As we can see from Fig. 12, the compensator that is wrapped in an aluminum shell has the same physical structure as the detector. So the only difference between them is the composition of the gas around the two detectors. The influence of external factors on test results could be reduced by such structure.

According to working mechanism of MOX gas sensor [27], heating the sensing material of MOX sensor with a heater can enhance the reversibility of the sensor and accelerate the response and recovery rates. The response and recovery time might be faster at higher temperature than room temperature, therefore, the performance of MOX gas sensor is affected by the heater temperature in gas sensor. But according to the tested results, the working environment temperature changes will has little effect on sensor performance for MOX gas sensors with one good temperature compensating structure.

The effect of humidity on the performance of MOX gas sensor has not been tested due to the lack of testing setup. According to literature [31], humidity effects maybe neglected as they are usually an order less significant than temperature effects which is due to temperature change in the sensor heating element

B. Field Tests of Leak Detection System

In order to verify our system, the experiment was carried out, as shown in Fig. 15. For safety, the volatile and non-toxic alcohol (95% alcohol purity) is selected as the source of gas. The source of the leak consists of air pump, power supply, alcohol bottle and nozzle. In the leak process, the gas concentration is constant. Hence, the leakage only depends on the size of the nozzle. In the experiments, the nozzle sizes are 6 mm, 11 mm, 20 mm. 30 leaks were created and each one lasted 3 minutes. It takes about five minutes to initialize the gas sensor before testing (sensor resistance preheating), and the node will work well after a preliminary classification and Zigbee restart needs time until become active, it's all about 10 seconds to start measuring in the system.

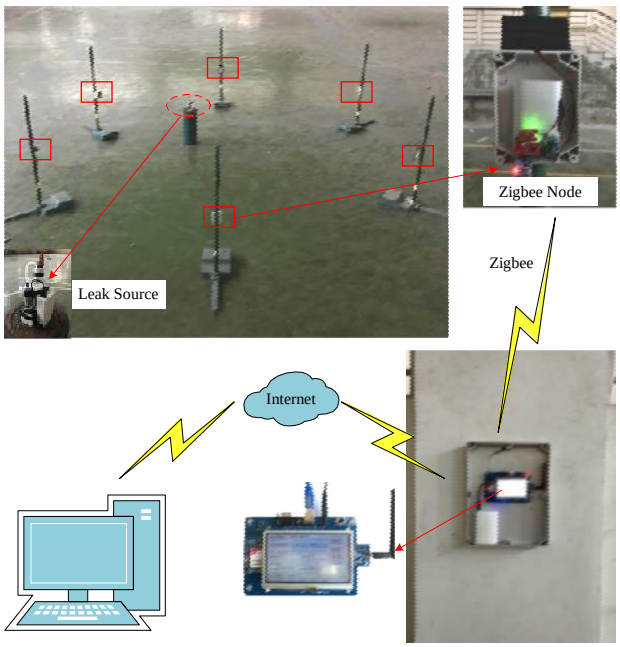
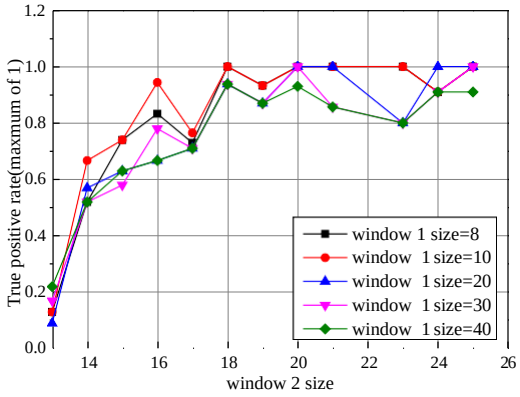


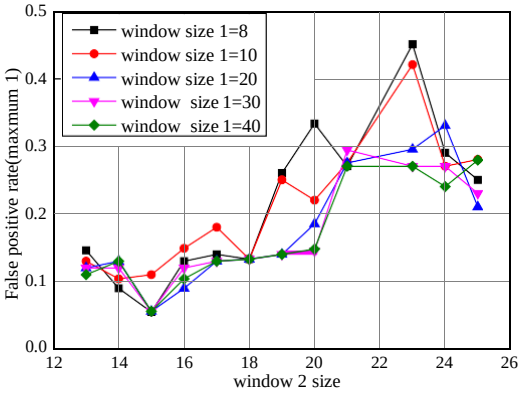
Fig. 15. Site of the experiment

C. Parameters Selection for Gas Leak Detection

Several parameters have significant effects on the judgment in the proposed method. They are analyzed respectively. As shown in Fig. 16 (a), the general trend observed show that the true positive rate of the detections increases as the length of the window 2 sizes increases greatly. And the impact of the size of window 2 on the true positive rate tends to be stable beyond a particular point (around 18). The length of window 1 shows little influence on the true positive rate of the detections. The impact of changing the window 1 size and window 2 size on the false positive rate is shown in Fig. 16 (b). It demonstrates that the false positive rate fluctuates at a lower level when the size of window 2 is small and then increases significantly when the length of window 2 exceeds a value (about 18). The trend shows that increasing the window 1 size does not affect the false positive rate obviously, except when it reduces to a small value (such as 8 or 10).



(a) The effect of the window size on the true positive rate.



(b) The effect of the window size on the false positive rate.

Fig. 16. The effect of the window size on both the true and false positive rate.

Besides the window sizes, the threshold is an important parameter that determines the performance of the detection method. The effect of the threshold on both the true and false positive rate can be seen in Fig. 17.

According the general trend, the true and false positive rate increase drastically. Furthermore, there are no detectable leaks and false positive judgment when the threshold is particularly small (set to 0). In addition, almost all of the segments will be misjudged (false positive rate of 100%) in another extreme case in which the threshold set to 1. The true positive rate is close to 100% at the same time because any degrees of the leak segments are certainly smaller than 1. However, false positive rate of 100% is unacceptable. Therefore, the threshold should be set to decrease the false positive rate on the premise of ensuring high true positive rate. In Fig. 17, the threshold is set at around 0.68 and the corresponding probability P_{sh} is 0.02.

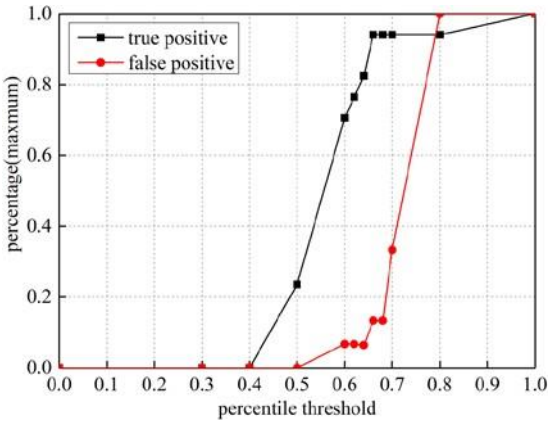


Fig. 17. The effect of the threshold on both the true and false positive rate.

D. Results

According to the parameters selection rules presented in subsection IV-C, the parameters are set as $NN = 18, MM_1 = 40$. In this test, historical data were collected for one day offline and more than four thousand segments were got as training samples when no leaks. Then the statistical distribution curve of degrees. After fitting, the PDF curve $PP(RR_{ii}|\theta\theta = 0)$ is obtained. So the threshold is set as 0.68 when the probability P_{sh} is 0.02. By using real-time data processing methods, real-time degrees are obtained and the results of leak detection are shown in the column of Table I. Based on the field operation and corresponding inspection records, during the monitoring of the set of 50 sets of leakage test group and 50 sets of non-leakage test group, the method of this paper successfully detected 47 sets of leakage data. When detecting the non-leakage data, 5 sets of false alarms occurred, and the delay was within 30S. Furthermore, the method of likelihood function [24] is used for comparison. A pipeline small leak detection by harmonic wavelet is used in the contrast experiment. The experimental setup of comparative approach is as follow:

- Stage 1 window size: 20;
- Stage 2 window size: 150;
- Percentile threshold: 15%.

The PDF in comparative approach is obtained based on the statistics distribution of the historical data in preparation stage. And the likelihood of sensor data is obtained. By dividing the segments, the similarity of the segments is calculated. The percentile threshold are determined by observing the general trend and keeping trying. The detection result are given in the Table I. There are 42 true positive alarms and 7 false alarms. It is obvious that the proposed method in this work is better than the comparative method in the term of detection efficiency. Moreover, the quite good performance is obtained in the detection delay. The time of the proposed approach is within 30s while the comparative approach is more than 100s averagely. In general, the method proposed by this study is more efficient.

TABLE I
THE COMPARISON OF TEST RESULTS BETWEEN COMPARATIVE AND PROPOSED METHODS

Signal type	Comparative method	Proposed method
Artificially leak signal	50	50
True positive alarm	42	47
Experimental group of no leakage	50	50
False positive alarm	7	5
Delay time	≥100	≤30

V. CONCLUSION

In this paper, a concentration-based leak detection method with statistical feature extraction is proposed. The detection model is built with the feature extracted from normal concentration signals. The approach obtained the non-concentration threshold using the distribution of the correlation between ACFs of normal concentration segments. Moreover, an automatic procedure is described to select sensor nodes that participate in leak detection considering the location of sensors. And the real-time concentration is the last segment in the real-time diagnosis process, which reduced greatly the detec- tion delay and calculation. Additionally, the analysis of the position of the sensors by the simulations of Fluent is performed in consideration of the influence of wind direction and obstacles during the deployment of the sensor network. Modeling simulation results provide recommendations for node installation. And the sensors are placed at the points where the concentration of gas is more easily collected. Based on the field operation (with 8 sensors and a monitoring area of 50 mm^2) and corresponding inspection records (with an average delay of 30 s), there were 47 true positive alarms,5 false positive alarms with the proposed method. The experimental results demonstrate that the proposed method in this work can effectively detect the gas leaks.

In the future, many topics remained for research and gas leak detection will make great progress. The noise in the original signals has great impact on the detection of small leaks and should be considered. With the improvement of wireless communication reliability, and as efficient microcontrollers become cheaper and less power consumption, the only component need to be improved is the sensor. The power consumption of gas sensors is a huge obstacle to the application of WSN in the field of gas leak detection and a gas sensors can only measure one gas; Detection of multiple gas types will result in increased energy consumption of nodes, which provides an opportunity for development of the energy harvesting and MEMS process. Better sensor hardware can provide more accurate data for algorithms.

REFERENCES

[1] State Administration of Work Safety, “Statistics of hazardous chemical accidents”, State Administration of Work Safety, retrieved on Dec. 2017[Online]. Available: http://www.chinasafety.gov.cn/newpage/wxhxp/wxhxp2013_jdys.htm

[2] Murvay P S, Silea I, “A survey on gas leak detection and localization techniques,” Journal of Loss Prevention in the Process Industries, vol. 25, no. 6, pp. 966-973, 2012.

[3] S P. Spachos, L. Song and D. Hatzinakos, "Gas leak detection and localization system," *2014 IEEE 11th Consumer Communications and Networking Conference (CCNC)*, Las Vegas, NV, 2014, pp. 1130-1131.

[4] S L. Shu, M. Mukherjee, X. Xu, K. Wang and X. Wu, "A Survey on Gas Leakage Source Detection and Boundary Tracking with Wireless Sensor Networks," in *IEEE Access*, vol. 4, pp. 1700-1715, 2016.

[5] W H. Wang, L. Dong, W. Wei, W. S. Zhao, K. Xu and G. Wang, "The WSN Monitoring System for Large Outdoor Advertising Boards Based on ZigBee and MEMS Sensor," in *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1314-1323, Feb.1, 1 2018.

[6] A. Somov, A. Baranov, D. Spirjaki, A. Spirjakin, “Deployment and evaluation of a wireless sensor network for methane leak detection” *Sensors & Actuators a Physical*, vol. 202, no. 10, pp. 217-225, 2013.

[7] V. Jelicic, M. Magno, D. Brunelli, G. Paci and L. Benini, "Con-text-Adaptive Multimodal Wireless Sensor Network for En- er-gy-Efficient Gas Monitoring," in IEEE Sensors Journal, vol. 13, no. 1, pp. 328-338, Jan. 2013.

[8] D. Brunelli, M. Rossi, "Enhancing lifetime of WSN for natural gas leakages detection", Microelectron. J., vol. 45, no. 12, pp. 1665-1670, 2014.

[9] Aliyu F, Al-Shaboti M, Garba Y, et al, “Hydrogen Sulfide (H2S) Gas Safety System for Oil Drilling Sites using Wireless Sensor Network” *Procedia Computer Science*, vol. 63, pp. 499-504, 2015.

[10] Zhao L, Yang H, “Small-target leak detection for a closed vessel via infrared image sequences,” *Infrared Physics & Technology*, vol. 81, pp. 109-116, 2017.

[11] Universidad Carlos III de Madrid, “SENSIA’s Gas Imaging System,” Universidad Carlos III de Madrid, 2015.

[12] J. Leis and D. Buttsworth, "A Robust Method for Tuning Photoacoustic Gas Detectors," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4338-4346, May 2018.

[13] J. Leis and D. Buttsworth, "A Temperature Compensation Technique for Near-Infrared Methane Gas Threshold Detection," in *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1813-1821, March 2016.

[14] Emerson, “GDU-Incus Ultrasonic Gas Leak Detector Emerson Process Management,” Emerson, 2014.

[15] Akouemo H N, Povinelli R J, “Probabilistic anomaly detection in natural gas time series data,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 948-956, 2016.

[16] P. Gupta, A. Goyal, J. Dauwels and A. Ukil, "Bayesian detection of leaks in gas distribution networks," *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, Florence, 2016, pp. 855-860.

[17] F. Wang, W. Lin, Z. Liu, S. Wu and X. Qiu, "Pipeline Leak Detection by Using Time-Domain Statistical Features," in *IEEE Sensors Journal*, vol. 17, no. 19, pp. 6431-6442, Oct.1, 1 2017.

[18] Xiao Q, Li J, Bai Z, et al, “A Small Leak Detection Method Based on VMD Adaptive De-Noising and Ambiguity Correlation Classification Intended for Natural Gas Pipelines,” *Sensors*, vol. 16, no. 12, pp. 2116, 2016.

[19] Y. Yingchun and Z. Laibin, "De-noising of acoustic signal by natural gas pipeline leakage based on DTCWT-SVD," *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)*, Hefei, 2016, pp. 1551-1556.

[20] Hu J, Zhang L, Wei L, “Detection of small leakage from long transportation pipeline with complex noise,” *Journal of Loss Prevention in the Process Industries*, vol. 24, no. 4, pp. 449-457, 2011.

[21] J. Kang, Y. J. Park, J. Lee, S. H. Wang and D. S. Eom, "Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems," in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4279-4289, May 2018.

[22] C. Varon, D. Testelmans, B. Buyse, J. A. K. Suykens and S. Van Huffel, "Robust artefact detection in long-term ECG recordings based on auto-correlation function similarity and percentile analysis," *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA, 2012, pp. 3151-3154.

[23] M. Lavielle, "Optimal segmentation of random processes," in *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1365-1373, May 1998.

[24] F. Chraim, Y. Bugra Erol and K. Pister, "Wireless Gas Leak Detection and Localization," IEEE Transactions on Industrial Informatics, vol. 12, no. 2, pp. 7*68-779, April 2016*.

[25] Kim, Y.H.; Kim, S.J.; Kim, Y.-J.; Shim, Y.-S.; Kim, S.Y.; Hong, B.H.; Jang, H.W. Self-activated transparent all-graphene gas sensor with endurance to humidity and mechanical bending. ACS Nano 2015, 9, 10453–10460.

[26] A. Somov, E.F. Karpov, E. Karpova, A. Suchkov, S. Mironov, A. Karelin, A. Baranov, D. Spirjakin, Compact low power wireless gas sensor node with thermo compensation for ubiquitous deployment, IEEE Transactions on Industrial Informatics, 2015,vol.11, no.6, pp.1660-1670.

[27] Hu Long, Anna Harley-Trochimczyk, Thang Pham, Zirong Tang, Tielin Shi, Alex Zettl, Carlo Carraro, Marcus A. Worsley, and Roya Maboudian, “High Surface Area MoS2/Graphene Hybrid Aerogel for Ultrasensitive NO2 Detection”, Advanced Functional Materials, 2016, pp.1-6.

[28] Ioannis Mandourarakis, Eftichios Koutroulis, Unified System- and Circuit-Level Optimization of RES-Based Power-Supply Systems for the Nodes of Wireless Sensor Networks, IEEE Transactions on Industrial Informatics, VOL. 14, NO. 2, FEBRUARY 2018, p598-607.

[29] Mingyi Chen, Min Wang, Hengwei Yu, Guanghui He, Yongxin Zhu, Yadong Liu, and Guoxing Wang, A Self-Powered 3.26- μ W 70-m Wireless Temperature Sensor Node for Power Grid Monitoring, IEEE Transactions on Industrial Electronics, Vol. 65, No. 11, November 2018, pp.8956-8965.

[30] F. Chraim and K. Pister, "Wireless valve position monitoring: A MEMS approach," *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Vienna, 2013, pp. 4016-4021.

[31] N. Masson, R. Piedrahita, M. Hannigan, Approach for quantification of metal oxide type semiconductor gas sensors used for ambient air quality monitoring, Sensors and Actuators B, 208 (2015), pp.339 – 345

International Conference on Latest Trends in Electronics and Communication (ICTEC) ISBN 978-93-88808-62-0

Influence Of Wind Power Plants On Power Systems Operation

Poushya Kandukuri 1 ***DRK Mahesh 2***

poushya99@gmail.com drkmahesh515@gmail.com

1,2 Assistant Professor , Malla Reddy College of engineering, Hyderabad,Telangana,India

Abstract— **In recent years, renewable energy sources (RES) has increased all over the world, especially in Europe, and the changes brought by these sources have a significant impact on system performance and efficiency. Integrating RES to an electric power network offers many techno-economical benefits and in the same time necessitate advances in the operation, control and planning of electric networks. This paper focuses on the influence of the interconnecting of wind power plants on the electric network in terms of optimal power system operation. In this paper is proposed a k-means clustering based approach to identify the typical load profiles for wind farms in order to achieve the optimal power system operation on minimizing the power losses.**

Keywords-wind power plant, optimal operation, typical load profiles, clustering techniques.

I. INTRODUCTION

The electricity industry restructuring, along with advances in small scale generation technologies, and a higher awareness of environmental issues are the key factors that have influenced the development of RES in the last period of time [1]. The electrical network of the future should be flexible, efficient, reliable and security of supply and in the same time will allow two way flows of energy and real time communication capable of self healing and enabling fast restoration from supply disruption facilitate market operations and customer choice and information [2].

The growing integration of the renewable energy sources like solar power, wind power, or combined heat and power in the power network impacts a lot of interested parties: transmission and distribution companies, the owners and operators of the distributed generation sources, other end users of the power network, regulators and policy makers.

An increasing introduction of RES without any changes in the electrical networks will ensue in unacceptable levels of quality and reliability. Small distributed generation sources (DG) are located to the medium or low voltage distribution network, where traditionally only consumption has been connected. The penetration of large quantities of them will require improvements not only at the voltage level where the sources are located but also at higher voltage levels. The intermittent character of renewable energy production introduces new power quality phenomena, representatively at lower voltage levels [3]. Besides, this variation as well as the difficulty in estimating RES production impacts the operation of the power transmission networks. These are the principal reasons why the RES integration is a very attractive research subject and it will continue being of great interest.

The changes brought by RES have definitely an important impact on system efficiency and performance and necessitate enhancements in operation and planning of electrical system. In order to maximize the potential RES benefits in the future could be taken enforcement measures or upgrading actions for improving the system performance and reliability. The upgrading process will be done, taking into consideration the electricity delivery infrastructure that is divided into transmission and distribution system, [4].

The traditional planning process for expanding transmission system is based mainly on its needs on past and designed loading levels, which have traditionally been assessed of future demand. In the deregulated market, and in the case of using different renewable energy sources, transmission planners must respond to the necessities of power generators. Otherwise, planning to develop transmission system may now be established by the location and type of generation source, rather than by the necessities of the transmission system.

Around the world, wind and photovoltaic power are considered to have the largest potential in electricity production from renewable energy sources, [5].

Currently wind energy has the largest interest because Romania has the highest potential from the Southeast of Europe, and Southeast of Romania ranks on the second place across the entire continent. In the past 10 years more studies considered necessary and appropriate addressing some activities of reassessment of wind potential of Romania, through the use of appropriate tools and instruments (measuring equipment, appropriate software) from measured wind data stations belonging to the National Agency of Meteorology, [6]. In [7] is given an estimate of the theoretical potential of wind energy corresponding (approximately 14 GW of installed capacity). Annual energy of this potential is 0,023 TWh / year.

Because in the last years there is an increase of wind farms connected to the grid powers, the electrical companies need by studies and analysis for to evaluate the impact of these sources on the system. These activities are performed in order to ensure reliable operation of the system in presence of wind farms [7], [8].

There are many requirements, prescriptions and guidelines that to refer to technical information necessary for assessment

of the influence on power system given by the wind sources. In the last years, there is a new approach that is based on the connecting to power systems of wind power farms with a large installed capacity (> 100 MW). Fig. 1 presents the influences of wind sources connected to the power systems, in function by the time horizon and dimension of surfaces that can be considered in the technical studies.

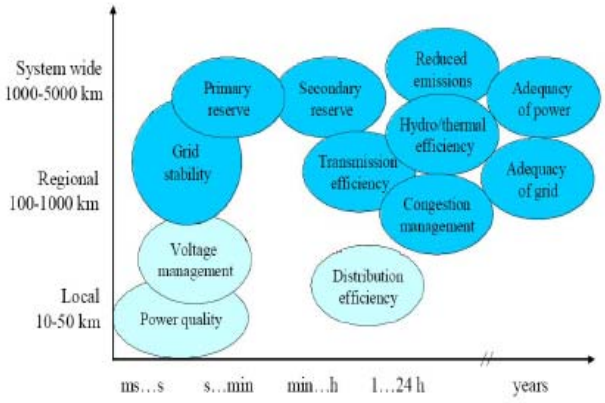


Figure 1. Influences of the wind sources connected to the power systems, [8]

The operation system of the wind turbine is conditioned by two parameters (speed wind - v and wind variations). There are three operating states that can be differentiated, [9], [10]:

Standstill of the turbine – for $v < v_{cut\text{-}in}$ or $v > v_{cut\text{-}out}$, where $v_{cut\text{-}in}$ and $v_{cut\text{-}out}$ represent cut-in and cut-out wind speeds .

Partial load – for $v_{cut\text{-}in} \leq v \leq v_n$, where v_n is the rated wind speed.

Full load – for $v_n < v \leq v_{cut\text{-}out}$. The value $v_{cut\text{-}out}$ is usually 25 m/s..

These three states represent characteristics P- v of the wind power plant. In order to study the influences of a wind sources on power system, it is needed to determine the patterns as accurately as possible. In the literature different techniques have been used for the classification and load profiling, but most of them were implemented to solve the problems from power systems. A review of the literature revealed two types of methods: statistical methods [11], [12] and methods based on artificial intelligence techniques fuzzy logic [13], [14], neural networks [15], data mining [16], clustering, [17]-[19].

In the paper an approach able to identify the optimal power systems operation considering the typical load profiles for wind farms is presented. First of all, an algorithm based on k-means method was used to find the typical load profiles for wind farms using a wind power generated database. After that, with this approach it can be find the optimal power system operation in order to minimize the power losses.

II. Determination Of Typical Load Profiles For Wind Power Plants

An approach based on clustering to determinate the wind power profiles for wind power farms from an electrical system is proposed. The *K*-means clustering algorithm used is used to classify operational profiles of wind power farms into coherent groups. By knowing these profiles, the operators can streamline the assessment of the demand.

The load profiling represents a different approach than the one based on metered demand. In this manner, for wind power farms from the electrical system are assigned a typical load profile. The shape of operational characteristic is influenced by the day (working or weekend) or season (spring, summer, autumn or winter). The operational characteristics of wind sources are in a very large number. This aspect create problems in analysis them. Thus, for an easy handling, they can be grouped in patterns, in function by the similarities between these. Every pattern will be characterized by a characteristic profile named typical load profile (TLP), [20], [21].

Each TLP is represented by a vector $x_i = \{x_{ih}, h = 1,\ldots, T\}$ for $i = 1, \ldots, K$, and the comprehensive set of TLPs is contained in the set $P = \{x_i, i = 1, \ldots, K\}$. The time scale along the day is partitioned into T time intervals of duration Δt_h , for $h = 1,\ldots, T$. Hourly values are used in this paper to exemplify the application. The variables used in the calculations are assumed to be represented as constant (average) values within each time interval. The clustering process forms K patterns corresponding the wind power plants. Further, the typical load profiles are assigned to wind power farms.

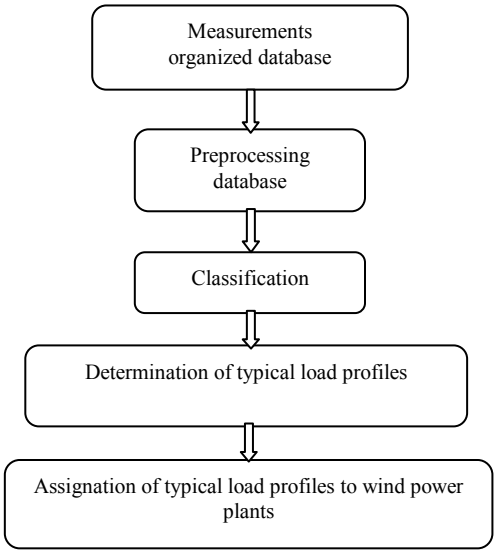


Figure 2. Flow-chart of the TPLs determination.

The proposed algorithm has the following steps, Fig. 2:

Step 1. Measurements: In this stage a database of operational characteristics is built.

Step 2. Data cleaning and pre-processing: A lot of technical aspects refer to communication problems, failure of equipment, etc., that can influenced negative the analysis need to be cleaned, pre-processed and reduced before the operational characteristics to be used in the clustering process.

Step 3. Classification: To realize this grouping, the K-means algorithm is used. Each operational characteristic is normalized. The normalization is made using the following relation:

$$z_h^{(j)} = \frac{x_h^{(j)}}{X^{(j)}} \quad ,j = 1, \ldots, N,\ h = 1, \ldots, T \qquad (1)$$

where:
$z_h^{(j)}$ – the normalized value;
$x_h^{(j)}$ – the measured value;
$X^{(j)}$ – the normalizing factor over the surveyed period (energy over analized period);
N – number of wind power farms from electrical system.

Step 4. Determination of typical load profile for wind farms: In this step, a refining of normalized characteristics occurs so that the unrepresentative characteristics are eliminated. Further, the TLP for each pattern is obtained using an averaging process of the hourly values. TLPs obtained can characterize very well the operation mode of the wind power farms, regarding to the electrical energy consumption.

5. Assignation: Finally, for each pattern of wind power farms, a TLP can be assigned.

III. Optimization Model

The objective of the planning formulation is to enhance the performance of the systems by minimize the active power losses. The main goal of the paper is to determine the optimal power system operation considering wind power sources in to an electrical network, minimizing the power losses.
A mathematical expression of the problem is:

$$\min F(X) = \min[P(U)]. \qquad (2)$$

where X is a power flow solution which stores data about the location in the system and the power capacity of the generators as well as of loads; $P(U)$ represents the power losses that depend of vector U.

To minimize the power losses into a electrical network, was used relationship (6), where R_{ij} is the resistance in to branch ij, i=1...n, j=1...n, $i \neq j$, P_i and Q_i are the real and reactive power into a node i, U_n the nominal voltage and $N_{branches}$ the number of branches in the network:

$$F_1 = \sum_{i=1}^{N_{branches}} \Delta P = \sum_{i=1}^{N_{branches}} \frac{R_{ij} * (P_i^2 + Q_i^2)}{U_n^2}. \qquad (3)$$

This item should compose with constraints to obtain the proper objective function. The main constraints in the process to determine the optimal power system operation with the proposed methodology are:

1. Voltage stability:

$$U_{i\min} \le U_i \le U_{i\max}. \qquad (4)$$

where: $U_{i\min}$, $U_{i\max}$ – minimum, maximum allowable voltage level in the system at bus i; U_i – voltage level at bus i.

2. Branch thermal limits – the power over the branch ij, S_{ij}, must be less than the maximum limit admissible that can support the line $S_{ij}^{\max}$.

$$S_{ij} \le S_{ij}^{\max} \qquad (5)$$

3. Generation limits - power generated by wind power plants is included between the maximum power allowed in bus i, $P_{DGi,\max}$.

$$P_{DGi} \le P_{DGi,\max}. \qquad (6)$$

4. Constraints for reactive power:

$$Q_{i\min} \le Q_i \le Q_{i\max}. \qquad (7)$$

where: $Q_{i\min}$, $Q_{i\max}$ - minimum, maximum allowable reactive power level in the system at bus i; Q_i - reactive power level at bus i.

5. The power losses after installing wind power plants in electrical network should be less than power losses before installing it.

$$\Delta P_{withDG} \le \Delta P_{withoutDG}. \qquad (8)$$

IV. Case Study

The proposed method was used for the optimal operation of a test electric network 220/110 kV with 10 nodes (3 nodes by 220 kV, and 7 nodes by 110 kV) and 12 branches (2, 220 kV electric lines; 7, 110 kV electric lines and 3, 220/110 kV power autotransformers), Fig. 3.

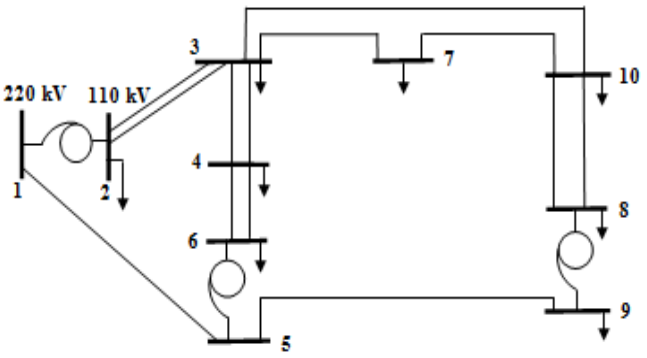


Figure 3. 220/110 kV test system.

In the test network a great potential based on wind energy was located in the region around the nodes 6 and 8. Thus, the wind power farm located in node 6 have a total installed capacity of 50 MW and the wind power farm placed in node 8, 25 MW.

In the case study a database described by generated power models, for the autumn season (3 months), corresponding to a group of wind power farms from the test electric network was considered. Every generated power model is described by 24 hourly points that depict the behavior of a wind power plant during a day.

The general information used in the clustering process concerns on generated power by the wind power plant, hour per hour for 3 months.

For determination of the optimal number of patterns, the algorithm presented in [21] was used. Thus, in the first step, the maximum number of patterns K_{max} must be determined. The

value of K_{max} was calculated with relation $K_{max} = \sqrt{N}$, where N represents the total number of characteristics from database ($N = 90$).

In the second step, the k-means clustering method with values for K between 2 and K_{max} is used.

In the step three, the quality of grouping is evaluated using the silhouette global (SG) coefficient, Fig. 4.

Figure 4. Variation of SG coefficient

From Fig. 4, it can observe that optimum value for K is 3. For this value, in Fig. 5 is represented the forms (silhouettes) of patterns.

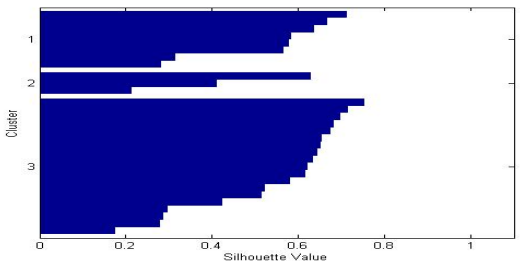


Figure 5. The forms (silhouettes) of patterns (case $K = 3$).

Each pattern is characterized by a TLP that was obtained by an averaging process of the hourly values. The wind generated power characteristics for each pattern are presented in Figs. 6-8. TLPs of the wind power farms corresponding to three obtained patterns (WPP1, WPP2, and WPP3) are indicated in Figs. 9-11.

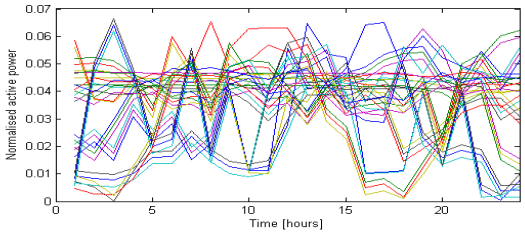


Figure 6. Generated power characteristics for WPP1pattern

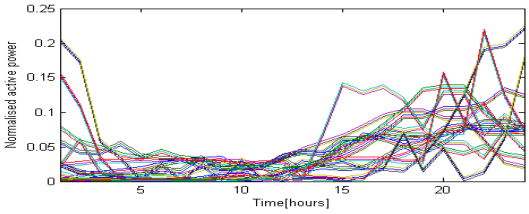


Figure 7. Generated power characteristics for WPP2 pattern

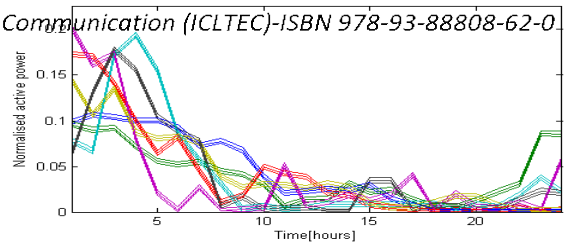


Figure 8. Generated power characteristics for WPP3 pattern

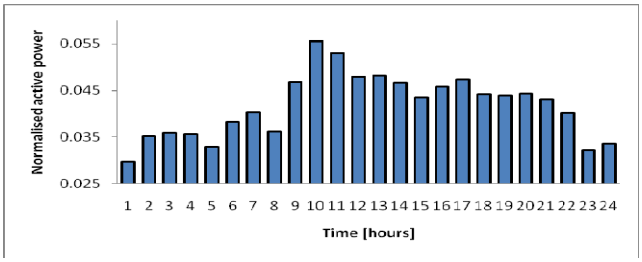


Figure 9. Typical load profile for WPP1 pattern

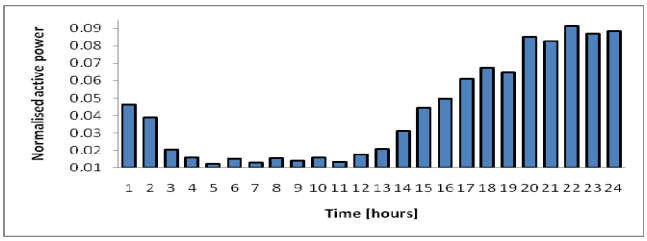


Figure 10. Typical load profile for WPP2 pattern

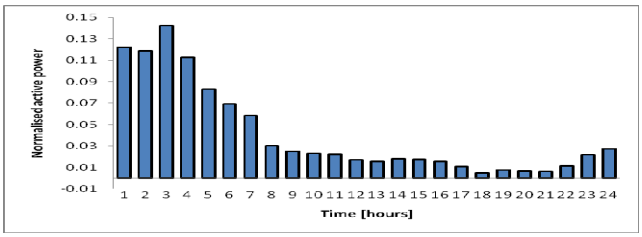


Figure 11. Typical load profile for WPP3 pattern

Further, considering these three TLPs are performed power flow calculations to analyze the evolution of the objective function. In this process the operating autotransformers plot was considered constant.

Fig. 12 shows the evolutions of the objective function in the 220/110 kV test network in the initial case (without wind power plants), in comparison with the cases when wind energy is injected coresponding to the three typical load profiles WPP1,WPP2 and WPP3.

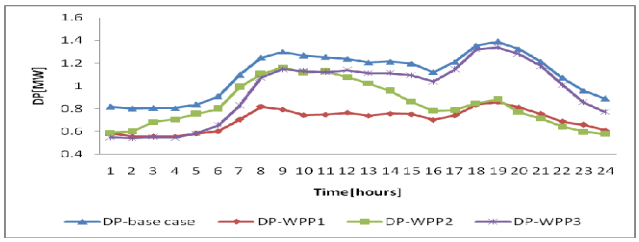


Figure 12. Objective function evolution in base case compared with cases WPP1, WPP2 and WPP3

Analyzing the results, it can see that a classification of the wind operation characteristics is useful to view the optimal operation and planning of an electric power system, on minimizing the power losses. The typical load profile WPP1 corresponds to the best power system operation for the test network, taking into consideration that has a full load operation state all the day. The objective function values, at peak load, in the base case and in case WPP1 and the voltage values evolution in nodes with wind energy injection are presented in Table I.

TABLE I. OBJECTIVE FUNCTION AT PEAK LOAD VERSUS VOLTAGE VALUES IN NODES WITH WIND ENERGY INJECTION

Objective Function				Voltage Values [kV]	
	DP[MW]	DP[%]		Base case	Case WPP1
DP-base case	1.3882	0.7744	Node 6	116.189	116.825
DP-WPP1	0.8584	0.4803	Node 8	113.674	114.297

CONCLUSIONS

In this paper a clustering technique based approach was proposed for determination the TLPs using a database described by wind power operation models, for the autumn season (3 months), corresponding to a group of wind power plants from the test electric network. The TLPs resulted describes very well the operation states of the wind power farms. So, WPP1 profile is characteristic for the operation at full load and WPP2 and WPP3 correspond to partial load operation state.

The results obtained demonstrate that the proposed approach can be used with success in the optimal power system operation on minimizing the power losses and in the same time to improve the voltage magnitude into an electric network.

REFERENCES

[1] H. Püttgen, PR. Macgregor and FC. Lambert, "Distributed generation: semantic hype or the dawn of a New Era?," IEEE Power Energy Mag., vol. 1, pp. 22-29, 2003.

[2] AG. Tsikalakis, ND. Hatziargyriou, "Environmental benefits of distributed generation with and without emissions trading," Energy Policy, vol. 35, No. 6, pp. 3395–3409, 2007.

[3] K. Alanne, A. Saari, "Distributed energy generation and sustainable development," Renewable & Sustainable Energy Reviews,vol. 10, pp. 539–558, 2006.

[4] V. Vittal, "The Impact of Renewable Resources on the Performance and Reliability of the Electricity Grid," Spring Issue of the Bridge on the Electricity Grid, vol. 40, no. 1, 2010. (https://www.nae.edu /Publications/Bridge/TheElectricityGrid/18587.aspx)

[5] A. Zervos, C. Lins and J. Muth, "A 100% Renewable Energy Vision for the European Union – Re-thinking 2050," http://www.rethinking 2050.eu/fileadmin/docum ents/ReThinking 2050_full_version_final.pdf, 2010.

[6] Renewable Energy Country Attractiveness Indices", Ernst & Young, 29 nov. 2012.

[7] Potențialul de energie regenerabilă în România, Gazeta de Agricultura, Categoria: Energie Regenerabila, http://gazetadeagricultura.info/ecobio/ 565-energie-regenerabila/11387-energie-regenerabila-in-romania.html, April 2011.

[8] Ö.S. Mutlu, "Evaluating The Impacts Of Wind Farms On Power System Operation," Journal of Naval Science and Engineering, vol. 6 , no.2, pp. 166-185, 2010.

[9] I.H. Yu, J. Lee, J.M. Ko, S.I. Kim, "A method for classification of electricity demands using load profile data," Proc.of the 4th Annual ACIS Int. Conf. Computer and Information Science, pp. 164 -168, 2005.

[10] H. Holttinen, P. Meibom, A. Orths, F. Hulle, B. Lange, M. O'Malley, J. Pierik, B. Ummels, J.O. Tande, A. Estanqueiro, M.Matos, E. Gomez, L. Söder, G. Strbac, A. Shakoor, R. João, J.C. Smith, M. Milligan and E. Ela, "Design and operation of power systems with large amounts of wind power," Final report, IEA WIND Task 25, 2009.

[11] Z. Lubosny, Wind Turbine Operation in Electric Power Systems. Berlin :Springer-Verlag, 2003.

[12] C.G. Carter-Brown, "Load profile modeling for integrated energy planning," Proc. of Domestic Use of Electrical Energy Conference, pp. 13-18, 1999.

[13] K.L. Lo, Z. Zakaria, "Electricity consumer classification using artificial intelligence," Proc. of the 39th Int. Universities Power Engineering Conf., Bristol, United Kingdom, 2004.

[14] D. Gerbec, S. Gašperič, I. Šmon, F. Gubina, "Determining the load profiles of consumers based on fuzzy logic and probability neural networks," in IEE Proc. Generation, Transmission and Distribution, vol. 151, pp. 95–400, 2004.

[15] M. Sarlak, T. Ebrahimi, S.S Karimi Madahi, "Enhancement the accuracy of daily and hourly short time load forecasting using neural network," Journal of Basic and Applied Scientific Research; vol. 2, no. 1, pp. 247-255, 2012.

[16] A.H. Nizar, "Load profiling and data mining techniques in electricity deregulated market," Power Engineering Society General Meeting, IEEE/PES, 2006.

[17] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," Energy, vol. 42, no. 1, pp. 68-80, 2012.

[18] N. Mahmoudi-Kohan, M.P. Moghaddam, S.M. Bidaki, "Evaluating performance of WFA K-means and Modified Follow the leader methods for clustering load curves," Proc. IEEE/PES Power System Conference and Exposition, 2009.

[19] Gh. Grigoras, M. Istrate, Fl. Scarlatache, "Electrical energy consumption estimation in water distribution systems using a clustering based method," 5th International Conference Electronics, Computers and Artificial Intelligence, Pitesti, Romania, 2013.

[20] L. R. Garcia-Escudero, A. Gordaliza, "A proposal for robust curve clustering," Journal of Classification, vol. 22, pp. 185-201, 2005.

[21] Fl. Scarlatache, Gh. Grigoraş, G. Chicco, G. Cârţină, "Using k-means clustering method in determination of the optimal placement of distributed generation sources in electrical distribution systems", International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), pp. 953-958, 2012.

Design & Implementation of Efficient Multiplier Using Fixed Width RPR

M. Shiva Kumar[1], *Asst. Prof, Dept of ECE, Malla Reddy College of Engineering*
A. Suresh Reddy[2], *Asst. Prof, Dept of ECE, Malla Reddy College of Engineering*

Abstract

A reliable low area efficient multiplier is designed in this paper by using algorithmic noise tolerant architecture. The ANT architecture can achieve the demand of low power, high precision and area efficiency. ANT architecture contains a main digital signal processing block along with an error correction block. In error correction block a replica of main DSP block with reduced operands named as RPR block is used for error detection. Here different multipliers are used in main DSP block to check which performs well in ANT Architecture. The multipliers used in this paper are Baugh Wooley multiplier, Wallace Tree multiplier, Row Bypassing multiplier and Bypassing multiplier (Row and Column Bypassing multiplier).

Key words: **ANT Architecture, RPR.**

I. Introduction

In VLSI lowering the power and area of systems is the main aim. To lower the power dissipation voltage over scaling is in the process because CMOS circuit power is directly proportional to the square of supply voltage. This voltage over scaling leads to the reduction of signal to noise ratio (SNR). In algorithmic noise tolerant architecture voltage over scaling can be used to reduce the power but signal to noise ratio is maintained because of error correction block. Previously full width RPR is used in error correction block but to reduce complexity full width RPR is replaced with fixed width RPR. Using the fixed width RPR computation error can be occurred. A compensation circuit must be added to the fixed width RPR to reduce the computation error. By taking the use of probability, statistics and partial product weight analysis, an approximate compensation vector is found.

II. ANT Architecture:

There are two blocks present in the ANT architecture. One is main digital signal processing block another one is error correction block. Error correction block contains a reduced precision replica block along with the circuitry which checks the error occurred because of voltage over scaling in main digital signal processing block. The ANT architecture is as shown in fig.1.

Here the output of main block is referred as $y_a[n]$, output of RPR block is referred as $y_r[n]$ and output of error correction block is noted as $\hat{Y}[n]$.

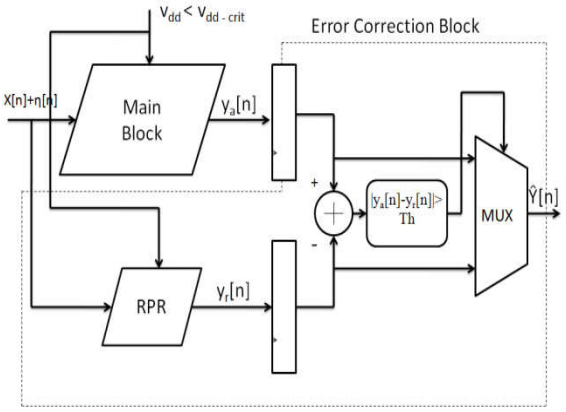


Fig1.ANT Architecture

RPR is a replica of main block with reduced precision operands and have shorter computation delay. Here fixed width RPR is taken instead of full width RPR to avoid infinite growth of bit width. If any soft errors occurred in main DSP block output $y_a[n]$, RPR output $y_r[n]$ is still correct since the critical path delay of replica is smaller than $T_{samp.}$ So $y_r[n]$ is used to detect errors in MDSP output by comparing the difference of $|y_a[n]-y_r[n]|$ against a threshold value Th. If the difference between $y_a[n]$ and $y_r[n]$ is larger than the threshold value Th, the output $\hat{Y}[n]$ is $y_r[n]$ else the output is $y_a[n]$. So $\hat{Y}[n]$ is expressed as

$$\hat{Y}[n]= y_a[n], \text{ if } |y_a[n]-y_r[n]| \leq Th$$

$$y_r[n], \text{ if } |y_a[n]-y_r[n]| > Th$$

The threshold value Th is determined as

$$Th= max_{\forall input} |y_0[n]-y_r[n]|$$

Where $y_o[n]$ is referred as error free output signal.

A full width (n/2) bit RPR can be divided into four subsets, which are most significant part (MSP), input correction vector (ICV), minor input correction vector (MICV) and least significant part (LSP). In the fixed width RPR only MSP part is kept and other parts are removed i.e. ICV, MICV, and LSP parts are truncated.

In the consideration of fixed width RPR a disadvantage is also present that there is a truncation error occurred because of the avoidance of ICV, MICV, LSB bits. To reduce the error, compensation circuit can be added to the fixed width RPR. The bits which are having highest weight in the truncated part are added as compensation circuit. Here ICV and MICV parts are used in compensation circuit because of their highest weighing.

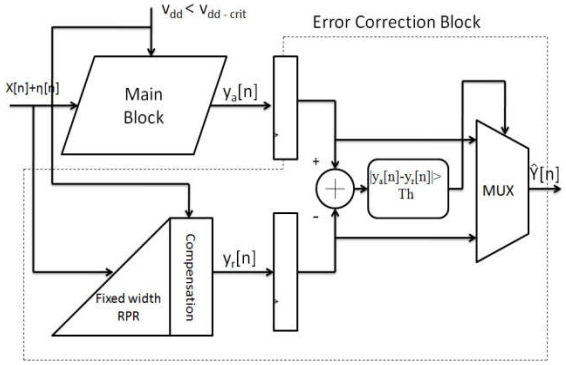


Fig 2: ANT Architecture with RPR compensation circuit

III. Baugh Wooley Multiplier

Baugh Wooley multiplier is used in the Main DSP block. Consider two unsigned inputs X and Y which are expressed as

$$X=\sum_{i=0}^{n-1} x_i\, 2^i \quad Y=\sum_{j=0}^{n-1} y_j\, 2^j$$

The result of multiplication can be expressed as

$$P=\sum_{k=0}^{2n-1} p_k\, 2^k=\sum_{j=0}^{n-1}\sum_{i=0}^{n-1} x_i\, y_j 2^{i+j} \qquad (1)$$

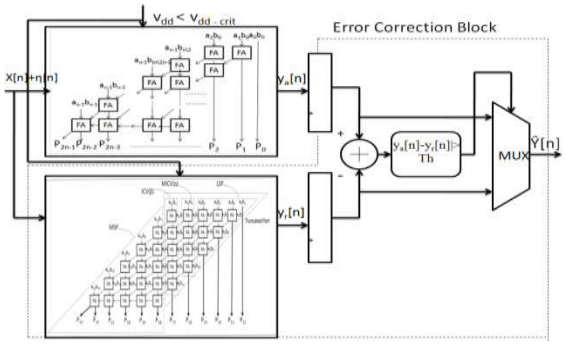


Fig 3: Baugh Wooley Multiplier in ANT architecture

For higher accuracy the error compensation circuit in the fixed width RPR can be set as shown in fig4.

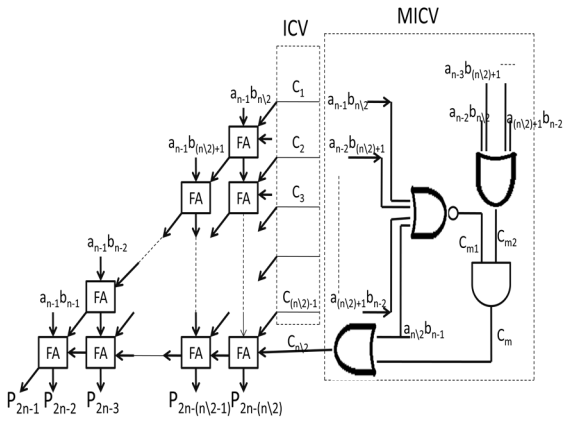


Fig 4: High accuracy fixed width RPR with compensation circuit constructed by ICV and MICV.

The output of fixed width multiplier P_t can be expressed as

$$P_t=\sum_{j=\left(\frac{n}{2}\right)+1}^{n-1}\sum_{i=\left(\frac{3n}{2}\right)-j}^{n-1} x_i\, y_j 2^{i+j} \quad + f(\mathrm{EC})$$

Where Error correction EC is expressed as

$$f(\mathrm{EC})=f(\mathrm{ICV})+f(\mathrm{MICV})$$

IV. Row Bypass Multiplier

Generally a conventional full adder has three inputs and two outputs. When the operand bit of multiplier is zero the full adder has disadvantages such as low operational speed and unwanted switching activity. When zero partial products are added, a large number of signal transitions are generated and do not affect the final product. By using Row Bypassing multiplier zero partial products can be bypassed to achieve optimization. A modified full adder is used to achieve this optimization. The modified full adder is as in fig 5.

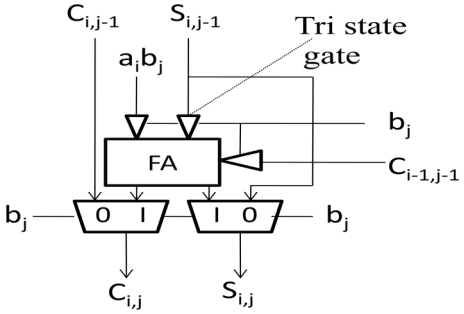


Fig 5: Structure of Full adder for Row Bypassing multiplier

In the design of n bit Row Bypass multiplier (n-1)x(n-1) full adders, 2x(n-1)x(n-1) multiplexers and 3x(n-1)x(n-1) three state gates are presented. A 4 bit Row Bypassing multiplier is as shown in fig 6.

A 12 bit Row Bypass multiplier can be designed and put in ANT architecture in the place of Baugh Wooley multiplier and performance was observed.

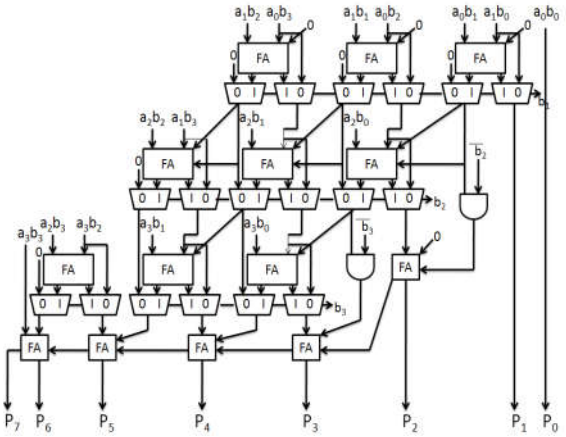


Fig 6: Row Bypass multiplier

V. Row and Column Bypassing Multiplier

Row and Column Bypassing multiplier is based on two dimensional bypassing features. According to this the addition operations in $(i+1)^{th}$ column or j^{th} row can be bypassed if the bit in that corresponding column or row is zero. The addition operation in the $(i+1,j)^{th}$ adder can be bypassed if the product a_ib_j is 0 and the carry bit $c_{i,j-1}$ is 0. If the product bit a_ib_j or the carry bit $c_{i,j-1}$ is 1, the addition operation in that corresponding full adder will be executed. It reduces the power more than that of row bypassing multiplier. The circuitry used to build this multiplier is changed as shown in fig 7(a) and 7(b).

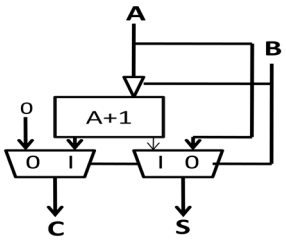


Fig 7(a): Half adder in Bypassing based design.

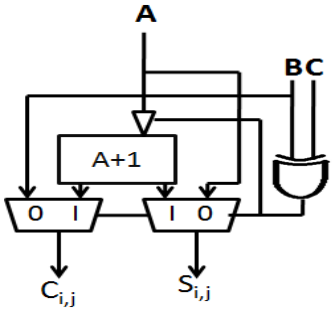


Fig 7(b): Full adder in Bypassing based design.

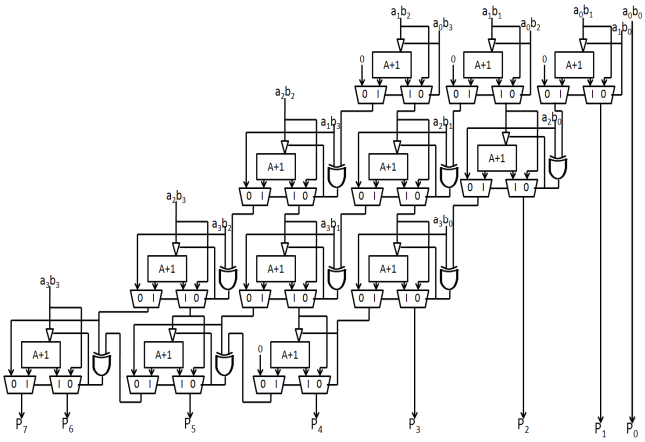


Fig 7(c): Row and Column Bypassing Multiplier

A 4 bit Row & Column bypassing multiplier is as shown in fig 7(c). A 12 X 12 bit Row & Column Bypassing multiplier is replaced in place of Baugh Wooley multiplier in the ANT architecture to check its performance.

VI. Simulation Results

These are the simulation results of ANT architecture when 12 bit Baugh Wooley multiplier, 12 bit Row Bypass multiplier and 12 bit Row and Column Bypassing multiplier are used.

Fig 8: simulation result of ANT architecture when Baugh Wooley multiplier is used.

Fig 9: simulation result of ANT architecture when Row Bypassing multiplier is used.

Fig 10: Simulation result of ANT architecture when Row and Column Bypassing multiplier is used.

VII. Conclusion

In the fixed width RPR based ANT design 12 bit Baugh Wooley multiplier, Row Bypassing multiplier and Row & Column bypassing multipliers are presented to check the area and delay performances.

The area and delay performance of Row & Column Bypassing multiplier is better than that of Baugh Wooley and Row Bypass multipliers.

The delay of Row & Column bypassing multiplieris 10% lower than Baugh Wooley multiplier and 7% less than that of Row Bypass multiplier. The area of Row & Column bypassing is 2% less than that of Baugh Wooley multiplier and 6% less than that of Row Bypass multiplier

References:

[1] I-chyn Wey, Chien-Chang Peng and Feng-Yu Liao " Reliable low power multiplier design using fixed width replica redundancy block " IEEE transactions on very large scale integration systems, vol.23, no.1, January 2015.

[2] B.Shim, S. Sridhara, and N.R.Shanbhag, "Reliable low power digital signal processing via reduced precision redundancy" IEEE Transactions on very large scale integration systems, vol.12, no.5, May 2004.

[3] Optimized Multiplier Using Bypass Technique" The International Journal of Computer Science & Applications, Volume1, no.2 April 2012.

[4] 'Low cost Low Power Improved Bypassing based Multiplier' A special issue on Recent Advances in Engineering & Technology 2013(NCRAET-2013).

[5] I.C.Wey and C.C.Wang, "Low error and area efficient fixed width multiplier by using minor input correction vector" in Proc. IEEE Int. Conf. Electron, Inf. Eng, vol.1, Kyoto, Japan, Aug 2010, pp.118-122.

[6] A.G.M.Strollo, N.Petra and D.D.Caro, "Dual tree error compensation for high performance fixed width multipliers" IEEE Trans. Circuits Syst II, Exp. Briefs, vol.52, no.8, pp 501-507 Aug 2005.

[7] "Area Efficient Low Error Compensation multiplier design using fixed width RPR" International Journal of Engineering and Technology e-ISSN:2395-0056 Volume: 02 Issue: 07 Oct-2015.

[8] "Design of low power fixed width multiplier using row bypassing" Article in IEICE Electronics Express, October 2012.

[9] J.M.Jou, S.R.Kuang and R.D.Chen "Design of low error fixed width multipliers for DSP applications" IEEE Trans, Circuits Syst, vol 46, no.6, pp. 836-842, Jun.1999.

[10] S.S.Kidambi, F.EI-Guibaly, and A.Antonou, "Area efficient multipliers for digital signal processing applications" IEEE Trans. Circuits Syst. II, Exp.Briefs, vol.43, no.2, pp.90-95, Feb. 1996.

[11] Y.C.Lim "Signal- precision multiplier with reduced circuit complexity for signal processing applications" IEEE Trans. Comput, vol.41, no.10, pp.1333-1336.

[12] "Reduction of Hardware Complexity and Truncation Error by using Fixed Width Baugh Wooley multiplier" Special Issue of International Journal of Computer Applications (0975-8887) on International Conference on Electronics, Communication and Information Systems (ICECI-12).

Modern Technology in Solar Energy Generation

Nimmala Mahesh[1],P. Manikya rao [2]
Dr. M.G.R Educational and Research Institute, [1,] Mallareddy college of Engineering[2]

Abstract:**Solar energy is playing a pivotal role in compensating the electrical energy as there is short fall in this engergy due to more demand and decline trends of conventional source of energies exhaustion of fuels like coal, petroleum, natural gases and constant of environmental and climatic changes to cope up this photovoltaic installation is being done in an electrical system to compensate and enhance the energy. an photovoltaic installation in an electrical system is made from the assembly of various photovoltaic units that uses solar energy to produce the electricity in a cheaper way from sun power. Till now the use and scope of solar energy is limited and has not reached upto masses Moreover the efficiency of the system is also low due to which the output is not sufficient as compared to input as in some installed case of solar panel it has been observed that efficiency is not more that 27%. To make it versatile and more useful for the masses newer trends and innovations will help. These have discussed in this paper.**

Keywords:**Solar panels, Efficiency, Namadd, renewable engergy resources, distributed generation**

I. Introduction

Now-a-days distributed generation (DG) is not a new concept. Without this the rectification of energy crises is not possible [1]. In most of the countries the electrical power demand is more than the electrical power generated. On the other hand there is a serious decline in the availability of natural resources, fuels, coal and gases etc. The generation of hydro power plant also varies due to variation in inflow of water from catchment area. When the capacity of hydro power plant decreases the power shortage arises. The solar power plant may be installed in such a fashion that these may work in unison for example when draught is more showering of sun is more. In this way shortage of power will be compensated by the energy governed by the solar power plant. Moreover this installation is to be done in such a way that solar panel will cover the rivers or reservoir reducing the evaporation which enhancing the capacity of dam. On the other hand the solar panel covering this area will generate electrical power which will the enhance the power generation of the system. Moreover by using some newer technologies this power generated may be integrated with the power grid to enhance the capacity of grid.

The development and uses of solar energy at large scale is not only reasonable method of energy resources utilization in the future but also effective frame to improving energy resource crises economically. There is different solar resource in different area, seasons, and weather conditions and so on because of so many influencing factors [2]. As this of kind of renewable energy is available in abundance in nature. The research and application of solar energy will be used to deal with alternative energy [3] – [4]. The advantages of renewable energy sources are enormous as they are free from gas emissions from few conventional energy resources which have impact on the global warming. If this generation of solar energy expedited rigorously can meet the most of the energy demand of the world. Use of Renewable energy will prove as panacea for solving the climatic and environmental problems as every sector of society is keen to solve these problems on the health ground problems. Currently, renewable energy sources install 15 percent to 20 percent of the world's total energy demand [5]. The solar energy is considered as the most promising and important renewable. It is envisaged that solar energy power plants would meet all human needs and would eventually replace the conventional power plants [6].

II. TYPES OF DISTRIBUTED GENERATION

There are different types of distributed generation according to the constructional and technical points of view as shown in Figure (1) [1].

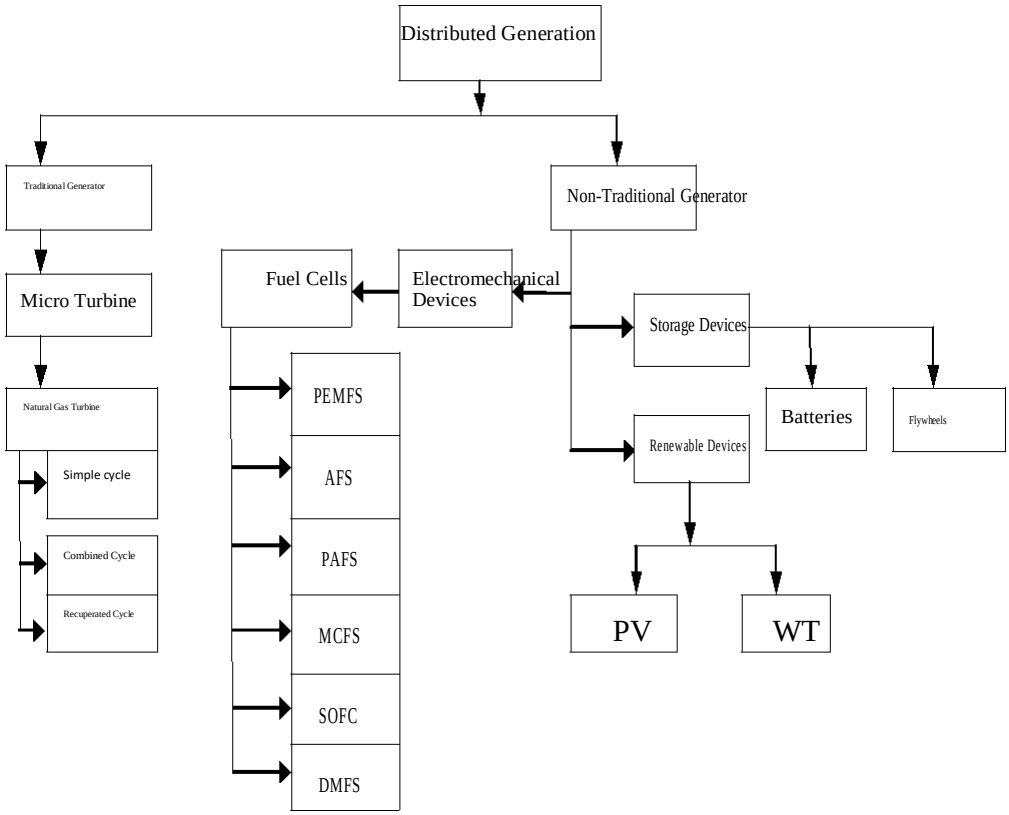


Figure 1 Types of Distributed Generation

III. The Principle Of Energy Storage

The circulation medium was heated by synchronous tracking and non-tracking solar collector and injected into the heat exchanger which was set in concrete pile with the help of circulation pump. After heat exchanging between the heat exchanger and concrete pile, circulation medium was pumped to synchronous tracking solar collector, so formed circulations.And then, the solar energy which was gathered by synchronous tracking and non-tracking solar collector was stored constantly in the underground concrete energy storage pile [7]. The circulating principle is shown in Figure(2). The solar nstal are used to generate steam which drives the steam turbines coupled with alternator as per figure (3).

IV. Solar Energy Policies Of India

The Government of India has increased its focus on developing alternativeresources of energies especially Solar Energy under the policies related to energydevelopment. The solar energy is available in abundance and almost free of cost as it is available from nature. Due to rapid economic expansion India is one of the

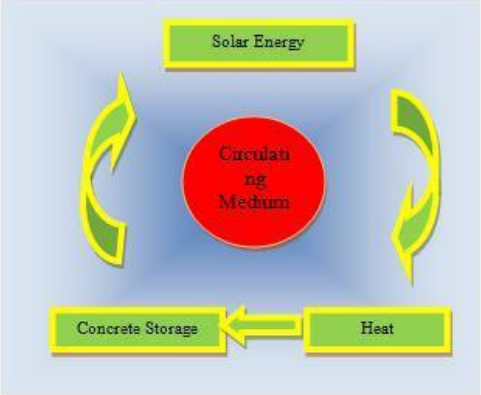


Figure 2 the circulating principle of solar energy

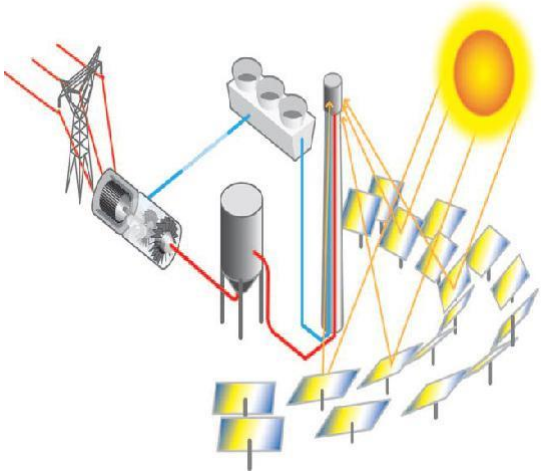


Figure 3 System of Power generation from solar energy

most growing markets and expected to be second largest energy contributor in energy market in the world by 2035. Due to limited domestic fossil fuels reserve, the India has strong planning to expand the renewable energy sources for power sector.

1. To supply the electricity to all the areas included the rural areas as mandated in section 6 of electricity act. Both the Central and State Government will jointly nstalled to achieve this objective at the earliest. Rural Electrification will be done for securing electricity access to the entire household in rural sector. Most of this requirement will be fulfilled by use of renewable energy sources.
2. Reliable rural electrification would be done either through conventional or non conventional methods of electricity whichever is more suitable and economical. Non conventional sources of energy especially Solar can be utilized even where Grid connectivity exists [8].
3. Particular attention is to be given to Dalit Bastis, Tribal areas and other weaker sections of the society the other newer resources.
4. Rural Electricity Corporation of India (REC) is the nodal agency at central govt. Level to implement these programs of electrification in rural areas. The REC will nstal all the goals set up by the National Common Minimum Programme ensuring timely implementation [9].
5. Responsibility of operation and maintenance & cost recovery could be discharged through appropriate arrangement with Panchayats, Local Authorities, BDO, and NGO etc [10].

6. This Great task of Rural Electrification requires cooperative efforts of all agencies like Govt. Of India, State Government and community education cell in rural areas.
7. The Electricity act 2003 has provision of restructuring the electricity industry which unbundled the vertically integrated electricity supply in each state. Now generation, transmission and distribution companies have been formed by the Regulatory Commission of state electricity board. Regulatory Commission will also specify the minimum percentage of electricity that each distribution utility must get from renewable energy sources [11].

V. Newer Methods Which Will Enhance the Use of Solar Energies

Day by day new trends and innovations are being developed throughout the world in R&D centers, automobile sectors and domestic use in institutions, hostels to reduce the energy wastage and to generate the power by solar devices. Many of them are explained given below:

1. Solar cells of higher efficiency have been developed having conversion efficiency more than 37% as compared to the previous solar cells having efficiency of 27 % made of two materials. Tata power is going to install the solar panels having 35% efficiency [12]. In These cells three photo absorption layers are stacked together. This has been developed by stacking Indium, Gallium and arsenide as the bottom layers. These cells have capability of absorbing the light from various wavelengths available in sunlight and convert into electrical energy. Through optimal process the active area has been increased. This breakthrough in technology has been done by new energy and industrial technology development organization.

1. The conversion efficiency of solar panels/plates is increased by newer devices of cleaning these panels. The device makes use of automated "dry-sweep'' to push dust and dirt away from the surface of these devices. In south Arabian language the device is known as nsta which is very rugged and have low maintenance cost. The device is powered by the lithium ion batteries. These batteries are charged by the array itself and have high efficiency. The device has moving parts. It is very interesting that this act like a robotic arm and automated work with scheduling. The device can jump the obstacle between the panels[13].

2. Throughout the world research and development is done to procure more and more energy from various devices and technologies. Under this concept Japan developed a fabric which is known as a solar cell fabric capable of harnessing the energy from sunlight while you are moving by wearing this fabric made cloth. This fabric is made from wafer thin solar cells woven in a stylish way. The electricity generated will be capable to charge the mobile and other portable electronic gadgets. The thread used will be stronger and which increases the life or durability of the fabric cloth. The same idea can be embedded/used in the blind makers and certain type of curtain will also generate power, when sun rayeson these. Various companies developing this type of fabric in association with solar cell maker. This will help the men to recharge these small gadgets while in sun [14].

3. A new trend of solar panel roofs have come in which most of the buildings the roof is covered with solar panel. In some advanced countries like china the roof of the max. Of the homes is made of solar panel by using aaluminum or strong alloy to support the weight of panels. In remote area where grid supply is not viable these panels generate power for themselves and supply electricity to the neighbours also who cannot afford the cost of installation. This will help nearby masses and community in that region where distribution of power is not feasible by other ways & transmission may not be possible due to heavy expenditure.

4. In some countries in urban area some hobbyist of solar energy generate electrical power this system for sufficient for their requirement and surplus generated power is supplied to the grid empowering the national grid. The solar panel should be installed on the vehicles where it is possible so that charging of batteries and other devices in the vehicle may be done with the help of solar energy. Whenever the solar rays fall on the panel this will improve the electrical efficiency of the vehicles. This type of experimentation and uses are already being done in Japan and in other advanced countries where conversion technologies from solar to electrical are being used frequently and sufficiently. In India also solar panels have been installed in metro railway service.

5. As we are aware of that electrical power demand is increasing and viewing the climatic concerns it is desired that renewable energy sources especially solar may be integrated to the utility grid. By using better flexibility in integration through power electronics. Harmonics can be reduced and the reactive power can be balanced.

6. In these days most of the power industries switchover side by side starting manufacturing of the solar inverters of high capacity. These may be utilized to get emergent power if not continues at remote locations where there is no grid supply.
7. CSP systems technology is used for power generation in the system large, flat, sunlight mirrors known as heliostats receive sun light at the top of the tower. A fluid for heat transfer is used to generation the steam which is used for production of electrical power .In some countries the capacity of these plant as high as up to 200 MW. These power tower are very popular in these days because of solar to electrical conversion efficiency is high [15].

VI. Distributed Solar Energy Generation

In [16] the authors has explained the use of Distributed solar photovoltaic (PV) systems is producing electricity onsite, so reducing the requirement to build up new transmission line and also avoiding line losses. Distributed generation also offer significant benefits to the consumers while providing resiliency to an electric grid that is based on the traditional and centralized model. These systems are used in applications ranging from small commercial to residential and for industrial use. Though this market is still primarily driven by government incentives, distributed solar PV will continue its steady march in future. Due to reduced market activity in Italy and Germany, global distributed solar photovoltaic market contracted slightly in 2012, However, growth in the United States, China, Japan, and other countries continued, driven by solar PV module price reductions, the growth of third-party financing models, and feed-in tariffs. Navigant Research forecasts that, from 2013 to 2018, 220 GW of distributed solar PV will be installed worldwide, representing $540.3 billion in revenue.

VII. Conclusion

Due to decline aviability of natural's fuels and viewing environmental changes causes due to conventional method of generation, the use ogf solar energy is becoming popular and urgency of the day. This will create healthy environment for the humain beings which are suffering from the various hazards due to pollution from the nstalled contents. Moreover the power generation due to hydro power plant is not also regular due to irrugalar flow of water from the catchment area. So it is concluded that solar power plant may be nstalled in such a way so these may work in unison with hydro and other methods of generation to enhance the clean and green energy.

References

[1] W El-Khattam and M.M.A Salama, "Distributed generation technologies, definitions and benefits", Electric Power Systems Research, vol. 71, Issue 2, p.p.,119–128, October 2004.
[2] Mingzhi Zhao, Zhizhang Liu and Mingjun Yu, "Data acquisition and analyzing of solar energy resource", Information and Automation (ICIA), Harbin, p.p., 2205-2208, 20-23 June 2010.
[3] Zhao Qingbo, Shan Baoguo, Situation and Outlook of Energy Demand in the World, Energy of China, 2002(02):34-36.
[4] Zehner and Ozzie, "Green Illusions", Lincoln and London: University of Nebraska Press. Pp., 331–42, 2012
[5] G.M. Shafiullah, A.M.T. Oo, D. Jarvis, .B.M.S. Ali, P. Wolfs, "Prospects of Solar Energy in Australia", Electrical and Computer Engineering (ICECE), 2010 International Conference, Dhaka p.p., 350 – 353, 18-20 Dec. 2010.
[6] S. Rao and B.B.Parulekar, "Energy Technology Non-Conventional, Renewable & Conventional", Khanna Publication, 3rd, 2012.
[7] Yan Zhao, Dajun Zhao, Yirong Yao, Xin Fang ,Shousheng Li, "Development of Solar Energy Underground Seasonal Storage Device and Its Parameters Measuring System", Information Engineering (ICIE), 2010 WASE International Conference, Beidaihe, Hebei, p.p., 231 – 234, 14-15 Aug. 2010.
[8] G. D. Kamalapur and R. Y. Udaykumar, "Rural Electrification in the Changing Paradigm of Power Sector Reforms in India", International Journal of Electrical and Computer Engineering, p.p., 147-154, vol. 2, No.2, April 2012.
[9] N. Sasidhar, "Electricity online trading in India", Rural Electricity Corporation of India, 2012.
[10] Giuseppe Buglione, Guido Cervigni, Eileen Fumagalli, Elena Fumagalli and Clara Poletti, "Integrating European Electricity Markets", Center for Research on energy and environmental economics policy, Report 2, Oct. 2009.
[11] MadhuKhanna, KusumMundra, AmanUllah, "Parametric and Semi-Parametric Estimation of the Effect of Firm Attributes on Efficiency: Electricity Generating Industry in India", Journal of International Trade and Economic Development. Pp. 419–436.vol. 8, No. 4, Sept. 2011.
[12] Electronics for You, "Most efficient Solar Cell", Technology news, January, 2013.

[13] GeorgeEitelhube, "Seed Fund Winners - NOMADD: The No-water Mechanical Automated Dusting Device", King Abdullah University of Science and Technology (KAUST), energy and sustainability are primary research drivers, 2012.
[14] John Badding, "Flexible Silicon Solar-Cell Fabrics May Soon Become Possible", journal Advanced Materials,Penn State University, 6 December 2012.
[15] DolfGielen, "Concentrating Solar Power technology", Energy Technology Systems Analysis Programme, International Renewable Energy Agency, January 2013.
[16] G. D. Rai, "Non-Conventional Energy Sources", Khanna Publication, 4th edition, 2012.

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

Talking without Talking: A Solution to Noisy Communication

K MANIKANTA
B.Tech M.Tech(ES & VLSI)
Department of ECE
Malla Reddy College Of Engineering
Hyderabad,INDIA

M.Rani
B.Tech,M.Tech(DECS)
Department of ECE
Malla Reddy College Of Engineering
Hyderabad,INDIA

***Abstract*— when we are in movie, theatre, bus, train there is lot of noise around us we can't speak properly on a mobile phone. In future this problem is eliminated with the help of Silent sound technology. It is a technology that helps you to transmit information without using your vocal cords. This technology notices every lip movements & transforms them into a computer generated sound that can be transmitted over a phone. Hence person on another end of phone receives the information in audio. It uses electromyography, monitoring tiny muscular movements that occur when we speak and converting them into electrical pulses that can then be turned into speech, without a sound uttered .When demonstrated, it seems to detect every lip movement and internally converts the electrical pulses into sounds signals and sends them neglecting all other surrounding noise. So, basically, it reads your lips. It is definitely going to be a good solution for those feeling annoyed when other speak loud over phone.**

***Keywords*-- silent sound, electromyography, electromyograms, digital image pr**

I. INTRODUCTION

Silent sound technology enables speech communication to take place when an audible acoustic sound is unavailable.

By acquiring sensor data from elements of the human speech production process- from the articulators ,their neural pathways, or the brain itself –it produces a digital representation of

speech which can be synthesized directly, interpreted as data, or routed into a communication networks[2].

Sound Technology is a technology for mobile phones that helps you communicate in noisy places too. It is a technology that will help reduce noise pollution to a great extent. The uses of this technology are immense for people who are vocally challenged or have been rendered mute due to accident.

Humans are capable of producing and understanding whispered speech in quiet environments at Silent remarkably low signal levels. Most people can also understand a few words which are unspoken, by lip-reading The idea of interpreting silent speech electronically or with a computer has been around for a long time, and was popularized in the 1968 Stanley Kubrick science- fiction film ''2001 – A Space Odyssey " [7].A major focal point was the DARPA Advanced Speech Encoding Program (ASE) of the early 2000's, which funded research on low bit rate speech synthesis ''with acceptable intelligibility, quality , and aural speaker recognizability in acoustically harsh environments".

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

Fig. 1 Many people talking at one place

II.PROCESS OF SPEAKING

The air passes through the larynx and the tongue and the words are formed with the help of the articulator muscles in the mouth and the jaw. The articulator muscles are activated irrespective of the fact that jo air passes through them or not. The weak signals are sent from the brain to the speech muscle. These signals are collectively known as the electromyograms.

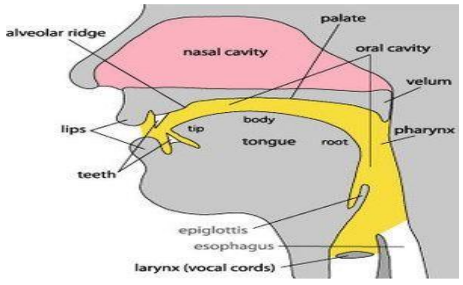


Fig. 2 Speaking Process in human body

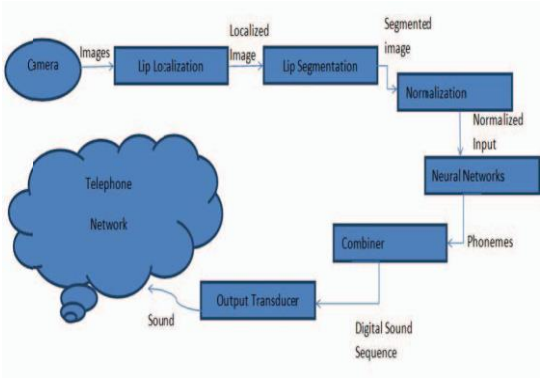


Fig. 3 Block Digram Showing different steps in generation of sound from lip movements

III.METHODS

Silent sound technology is processed in two ways

.They are

A. Electromyography (EMG)
B. Image Processing

A.Electromyography

Electromyography is a technique used in silent sound technology that monitors tiny muscular movements that occur when we speak and converting them into electrical pulses that can then be turned into speech, without a sound utter. Electromyography (EMG) is a technique for evaluating and recording the electrical activity produced by skeletal muscles.EMG is performed using instrument called an **electromyograph**, to produce a record called an **electromyogram**. An electromyograph detects the electrical potential generated by muscle cells when these cells are electrically or neurologically activated [2].

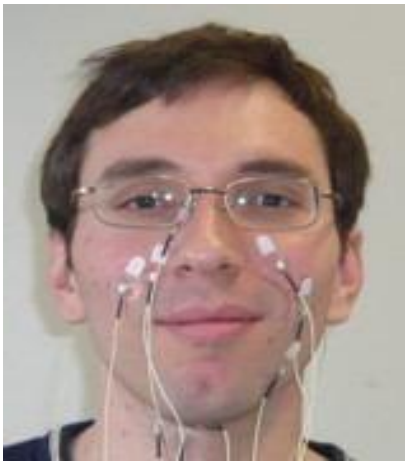


Fig. 4 Electromyographic sensors attached to face.

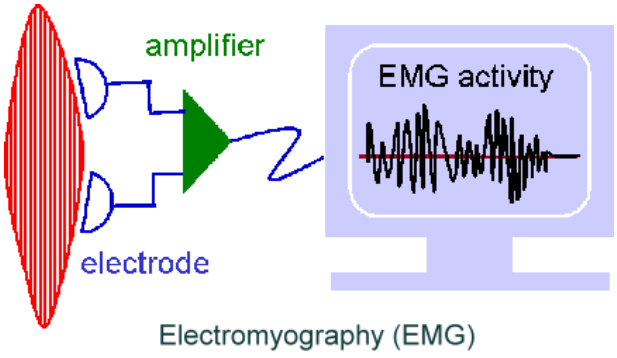


Fig. 5 Electromyography activity

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

B.Image Processing

The simplest form of digital image processing converts the digital data tape into a film image with minimal corrections and calibrations. Then large mainframe computers are employed for sophisticated interactive manipulation of the data. In the present context, overhead prospective are employed to analyze the picture. In electrical engineering and computer science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or, a set of characteristics or parameters related to the image. Most image- processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it [1].

Analysis of remotely sensed data is done using various image processing techniques and methods that includes

1) Analog Image processing: - Analog processing technique is applied to hard copy data such as photographs or printouts. It adopts certain elements of interpretation, such as primary element, spatial arrangement etc. With the combination of multi-concept of examining remotely sensed data it allows us to make a verdict not only as to what an object is but also its importance. Apart from these it also includes optical photogrammetric techniques allowing for precise measurement of the height, width, location, etc. of an object

2)

2)Digital Image Processing: - Digital Image Processing involves a collection of techniques for the manipulation of digital images by computers. It contain some flaws. To overcome the flaws and deficiencies in order to get the originality of the data, it needs to undergo several steps of processing. Digital Image Processing undergoes three general steps: 1) Pre-processing Display 2) Enhancement 3) Information extraction

IV. ADVANTAGES OF SILENT SOUND TECHNOLOGY

- Very useful for those people who lost their voice and has been rendered mute due to accident.
- At public crowded places like in market, bus, train, malls, theater etc.
- Very good technology for noise cancellation technique.
- Helps in making phone calls in noisy environment.
- Very useful for sharing confidential information like
- secret PIN number on phone at public place.

V.RESTRICTIONS

This technology works in many languages of user's choice like English, French & German, etc. But, for the languages like Chinese is difficult because different tones can hold many different meanings.

VI.FUTURE SCOPE

Silent sound technology gives way to a bright future to speech recognition technology from simple voice commands to memorandum dictated over the phone all this is fairly possible in noisy public places. Without having electrodes hanging all around your face, these electrodes will be incorporated into cell phones. Nano technology will be a mentionable step towards making the device handy.

VII.CONCLUSIONS

Engineers claim that the device is working with 99 percent efficiency. Silent Sound Technology, one of the recent trends in the field of information technology implements 'Talking Without Talking'. It will be one of the innovation and useful technology and in mere future this technology will be use in our day to day life.

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

REFERENCES

[1] Hueber T, Benaroya E-L, Chollet G, Denby B, Dreyfus G, Stone M. (2010). Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. Speech Communication.

[2] Denby B, Schultz T, Honda K, Hueber T, Gilbert J.M., Brumberg J.S. (2010). Silent speech interfaces. Speech Communication

[3] http://www.techpark.net/2010/03/04/ silent-soundtechnology-an-end-to-noisy-communications/

[4] Shehjar Safaya , Kameshwar Sharma , Silent Sound Technology-An End to Noisy Communication, Speech Communication , Vol. 1, Issue 9, November 2013

Delay Estimation Model for High speed Interconnects in VLSI

International Conference on Latest Trends in Electronics and Communication (ICTEC 2013) ISBN 978-93-82208-62-0

B.Balaji[1], Professor, Dept of ECE, Mallareddy College of Engineering
R. Purushottam Naik[2,] Professor,Dept of ECE, Mallareddy College of Engineering

Abstract: In recent days there is huge demand for high speed VLSI networks. In order to judge the behavior of on-chip interconnects the coupling capacitances and interconnect delays plays a major role. As we switch to lower technology there is on-chip inductance effect that leads to interconnect delay. In this paper we try to apply second order transfer function designed with finite difference equation and Laplace transform at the source and load termination ends. Analysis shows that current mode signaling in VLSI interconnects provides times better delay performance than voltage mode.

Keywords: Current Mode, Voltage Mode, VLSI Interconnect

I. INTRODUCTION

As the number of transistors on a chip continues to increase, on-chip communication becomes a more important facet of architectural design. Traditional electrical wires, typically driven by digital components using simplistic digital signals have issues to address in the scaling chip multiprocessor market, specifically latency and energy. Global wire latency remains relatively constant, translating to a larger relative latency for even moderately-sized systems. In order to ensure signal quality, digital repeaters and packet-switching routers must be added to facilitate the transmission of long distance communications, contributing further to the latency and energy issues. Current research focuses on a few categories of solutions, each with unique benefits and limitations. The current convention is the use of packet-switching networks topologies to provide the interconnect backbone for chip-multiprocessors. A packet-switched network-on-chip (NoC) provides in-field scalability, the ability to use commercial-off-the-shelf components, and high aggregate throughput. However, a NoC also requires higher power routers and potentially long latencies for long distance communication. Another state-of-the-art solution for interconnects uses on-chip optics. Research is currently being proposed to use either waveguides or free-space optics to provide a high-throughput, low-energy, low-latency medium for on-chip communication. On the other hand, optics also has issues that prohibit its immediate use as an interconnect backbone.

First, current optical components are not easy to integrate into standard silicon CMOS process, making it more difficult to fabricate with the current technologies without sacrificing electro-optical conversion efficiency. Additionally, while some on-chip lasers exist, most often, off-chip lasers are used to provide the optical power, shifting the onus of energy efficient operation off-chip, but not removing it from the system.

II. RELATED WORKS

Signaling in global strains is a main bottleneck in excessive performance VLSI systems because of the dominant problem of signal propagation delays in comparison to circuit delays. So, correct and accurate delay estimation models are required in cutting-edge-mode signaling. Accurate

estimation of propagation delay in worldwide interconnects plays a main place in the early design levels of VLSI systems compared to neighborhood interconnect delays because, global wires supports predominant capabilities like clock, sign distribution among the useful blocks and affords power/ground to all functions on a chip. Starting from lumped RC model to distributed RLC version, diverse techniques [6]-[8] based on analytical closed form formulations had been proposed to model delay in voltage-mode interconnects. Similarly for modern-day-mode RC interconnects, closed-form delay analysis version becomes presented in [9] and the evaluation does not encompass the fast aspect input reaction. In [10] closed-form delay model for allotted contemporary-mode RC line is provided, which included the practical rapid edge input reaction issue but this version should now not include the inductance impact in present day-mode interconnect. A delay estimation model [11], that's derived the use of the concept of soaking up inductance impact into equal RC version, then changed nodal evaluation (MNA) become used. Various closed-form delay models [9]-[13] for on-chip interconnects in present day-mode signaling have more inaccuracy in phrases of postpone estimation.

III. FINITE DIFFERENCE EQUATION AND LAPLACE TRANSFORM

The Taylor series expansion for a function of one variable about the point x is

$$f(x+h) = f(x) + h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

…….. (1)

The notation $O(h^3)$ indicates that the series, when truncated at the quadratic term in h, contains errors that scale as h^3 and higher powers of h.

We can immediately obtain an approximation to the derivative of f(x) from the first two terms of the expansion.

$$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h^1) \text{ …….. (2)}$$

Note that, even though we neglect terms of $O(h^2)$ in the expansion, since we divide through by h to obtain the derivative expression the approximation is correct to $O(h^1)$ only. Note that there is an asymmetry in this approximation to the derivative at x, since the function at x and x+h occur, but not the function at x-h. This is therefore referred to as a forward difference approximation. It is possible to expand f(x) in the negative direction in the Taylor expansion and hence to obtain a backward difference approximation

$$f(x-h) = f(x) - h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

…….. (3)

$$f'(x) = \frac{f(x) - f(x-h)}{h} + O(h^1) \text{ …….. (4)}$$

By combining forward and backward difference approximations it is possible to obtain a central difference approximation to the derivative of f(x) at x that contains errors of order $O(h^2)$. Note the difference in scaling of errors for the central difference approximation when compared with the forward and backward difference approximations.

$$f(x+h) = f(x) + h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

$$f(x-h) = f(x) - h\,f'(x) + \frac{h^2}{2!}f''(x) + O(h^3)$$

$$f(x+h) - f(x-h) = 2h f'(x) + 2\frac{h^3}{3!} f'''(x) + O(h^4)$$

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + O(h^2)$$

…….. (5)

By retaining terms in the Taylor series to order h^3 we can obtain an approximation for the second derivative which contains errors of order h^2

$$f(x+h) = f(x) + h f'(x) + \frac{h^2}{2!} f''(x) + \frac{h^3}{3!} f'''(x) + O(h^4)$$

$$f(x-h) = f(x) - h f'(x) + \frac{h^2}{2!} f''(x) - \frac{h^3}{3!} f'''(x) + O(h^4)$$

…….. (6)

Add these expansions to obtain

$$h^2 f''(x) = f(x+h) - 2f(x) + f(x-h) + O(h^4)$$

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + O(h^2)$$

…….. (7)

A. Taylor series expansions in more than one dimension

A PDE contains at least two independent variables and so we need to approximate differential operators in at least two dimensions. This is done using Taylor series expansions in more than one dimension. Suppose

u = u(x,y) …….. (8)

The Taylor series expansion of u about the point (x,y) is

$$u(x+h, y+k) = u(x, y) + h\, u_x(x, y) + k\, u_y(x, y) + \frac{h^2}{2!} u_{xx}(x, y) + \frac{2hk}{2!} u_{xy}(x, y) + \frac{k^2}{2!} u_{yy}(x, y) + O(|\mathbf{h}|^3)$$

$$\mathbf{h} = \begin{pmatrix} h \\ k \end{pmatrix}$$

…….. (9)

The vector notation for this expansion is

$$u(\mathbf{x}+\mathbf{h}) = u(\mathbf{x}) + \mathbf{h}^{\mathbf{T}} . \nabla u(\mathbf{x}) + \frac{1}{2!} \mathbf{h}^{\mathbf{T}} . \nabla\nabla u(\mathbf{x}) . \mathbf{h} + O(|\mathbf{h}|^3)$$

$$\mathbf{h} = \begin{pmatrix} h \\ k \end{pmatrix} \quad \mathbf{h}^{\mathbf{T}} =$$

$$\begin{pmatrix} h & k \end{pmatrix} \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix} \quad \nabla\nabla = \frac{\partial^2}{\partial x_i \partial x_j}$$

…….. (10)

Returning to the long-hand notation, the expansion of u(x,y) in the x direction is

$$u(x+h, y) = u(x, y) + h\, u_x(x, y) + \frac{h^2}{2!} u_{xx}(x, y) + O(h^3)$$

$$u(x-h, y) = u(x, y) - h\, u_x(x, y) + \frac{h^2}{2!} u_{xx}(x, y) + O(h^3)$$

…….. (11)

If we subtract these two equations and rearrange to make u_x the subject of the equation we find that the central difference approximation to u_x is

$$u_x(x, y) = \frac{u(x+h, y) - u(x-h, y)}{2h} + O(h^2)$$

…….. (12)

We can also obtain forward and backward difference approximations which contain errors of order $O(h^1)$ from the Taylor series expansion in either direction. It is convenient to write the function u at points on a grid for numerical solution with subscripted indices rather than arguments. Thus we make the equivalence

$$u(x, y+k) = u_{i+1, j} \quad \text{…….. (13)}$$

k is the step size or distance between gridpoints in the y direction in the numerical solution. Perversely, Farlow reverses the order of the arguments/indices in going over to the gridpoint index. We will do the same to maintain consistency with Farlow. The indicial representation of the first and second order partial derivatives is given below and illustrated

by corresponding 'computational molecules'.

$$u_x(x, y) = \frac{1}{2h}\left(u_{i,j+1} - u_{i,j-1}\right) + O(h^2)$$

$$u_y(x, y) = \frac{1}{2k}\left(u_{i+1,j} - u_{i-1,j}\right) + O(k^2)$$

$$u_{xx}(x, y) = \frac{1}{h^2}\left(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}\right) + O(h^2)$$

$$u_{yy}(x, y) = \frac{1}{k^2}\left(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}\right) + O(k^2)$$

…….. (14)

Provided that the step sizes in the two directions (h and k) are equal, then we obtain the following approximation for the Laplacian operator in 2 dimensions.

$$\nabla^2 u(x, y) = \frac{1}{h^2}\left(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}\right) + O(h^2)$$

…….. (15)

Voltage mode interconnects

Voltage mode signaling is most widely used in VLSI chips. In voltage mode signaling, receiver provides high input impedance (ideally infinity). The information is conveyed in the form of voltage. The output voltage is a function of input signal and is varied according to supply voltage. Fig.1 shows the theoretical model of conventional voltage mode interconnect implementation [5]. The output is terminated by an open circuit.

CMOS representation of voltage mode is shown in Fig. 2 [1, 3]. The driver consists of an inverter which drives long RC interconnect chain. This is terminated by high input impedance of the inverter circuit at the receiver. This high input impedance of the receiver gives rise to high input capacitance which leads to high charging and discharging time for RC interconnect chain. Hence voltage mode signaling has large delay. Due to high input impedance at the receiver, the charge accumulated at the input of the receiver does not get effective discharge path to ground as a result this may cause electrostatic induced gate oxide break down.

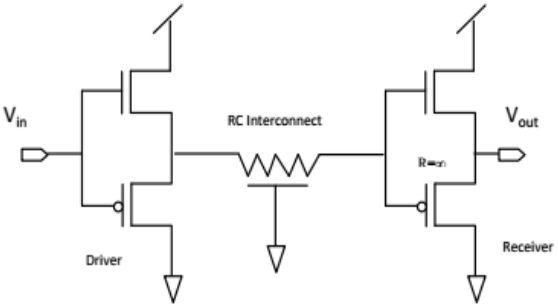


Fig. 2. CMOS representation of voltage mode signaling [5].

Current mode interconnects

In current mode signaling, information is represented as current signal. The receiver provides low impedance (ideally zero) at its input. In current mode signaling line is terminated by shorting the wire. The theoretical model of current mode signaling is as shown in Fig. 3 [5]

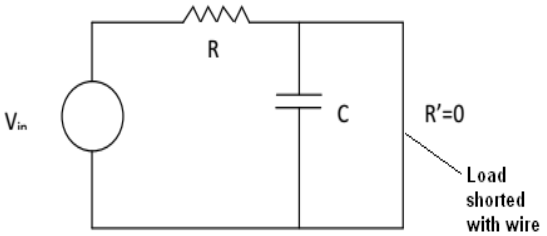


Fig. 3. Current mode signaling

The CMOS representation of current mode signaling is shown in Fig. 4. The receiver senses current signal at its input and provides low impedance.

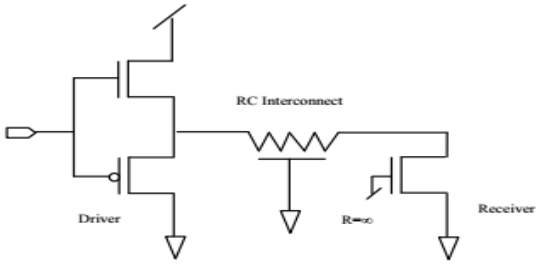
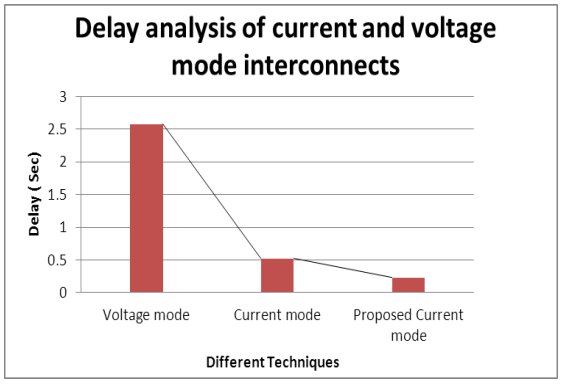


Fig. 4. CMOS representation of current mode signaling [5]

Table I Delay Analysis for voltage and Current Mode Interconnects [13, 14].

Signaling mode	Delay (ns)	P Total) (µw)
Voltage mode	2.58	22.01
Current mode	0.520	115.08
Proposed Current mode	0.010	19.33

Table I shows the delay analysis for both voltage and current mode interconnects [13-14]. It is analyzed that with Proposed Current mode interconnects delay decreases from 2.58ns to 0.010ns however power dissipation in the circuit increases from 22.01µw to 19.33µw. This is due to the low impedance at the receiver of current mode interconnect circuit. The overall figure of merit power-delay-product (PDP) of current mode increases. This displays the improved performance and advantage of current mode over voltage mode interconnects. The delay analysis for current and voltage mode interconnect is shown in Fig. 10. It is seen that there is 79.84% reduction in current mode interconnect delay.



IV. CONCLUSION

This paper presents second order transfer function designed with finite difference equation and Laplace transform at the source and load termination ends current mode signaling for delay estimation of current mode high speed VLSI interconnects. The perseverance of the current study is to estimate the delay of current-mode VLSI interconnects and to find the interaction between delay for various lengths, line inductances and load capacitances using existing voltage mode. All the benefits give current mode signaling an upper edge over the voltage mode signaling. At highly miniaturized technologies, interconnects with current mode signaling would be the best choice with the assistance of HSPICE tool.

REFERENCES

[1] V. Adler, and E.G Friedman, "Repeater design to reduce delay and power in resistive interconnect," IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 45, pp. 607-616, 1998.

[2] V. V.Deodhar, and J. A. Davis, "Voltage scaling, wire sizing and repeater insertion design rules for wave-pipelined VLSI global interconnect circuits," Sixth International Symposium on Quality of Electronic Design, pp. 592-597, 2005.

[3] R. Chandel, S. Sarkar, and R.P. Agarwal, "An Analysis of Interconnect Delay Minimization by Low-Voltage Repeater Insertion," Microelectronics Journal, Elsevier Science, vol. 38, no. 4-5, pp. 649-655, April-May 2007.

[4] R. Dhiman, and R. Chandel, "Sub Threshold Delay and Power Analysis of Complementary Metal-Oxide Semiconductor Buffer Inserted Interconnect for ultra-Low Power Applications," Journal of Low Power Electronics, ASP, vol. 8, no.1, 2012.

[5] A. Maheshwari, and W. Burleson, "Current sensing Technique for Global Interconnects in Very Deep Submicron (VDSM) CMOS," IEEE Computer Society Workshop on VLSI, pp. 66-70, 2001.

[6] W. C. Elmore, "The transient response of damped linear networks with particular regard to wideband amplifiers,"J. Appl. Phys., vol. 19, no.1, pp. 55–63, Jan. 1948.

[7] T.Sakurai, "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSI's,"IEEE Trans. Electron Devices, vol. 40, no. 1, pp. 118–124, Jan. 1993.

[8] Kahng A B, Muddu S, "An analytical delay for RLC interconnects", IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems, 1997, 16(12): 1507−1514.

[9] E. Seevinck, P. van Beers, H. Ontrop, "Current-Mohe Techniques for High-speed VLSI Circuits with Application to Current Sense Amplifier for CMOS SRAM's," IEEE J. of Solid-state Circuits, Vol 26, No. 4, April1991.

[10] Bashirullah, R.; Wentai Liu; Cavin, R., III, "Delay and power model for Current-mode signaling in deep submicron global interconnects," Custom Integrated Circuits Conference, 2002. Proceedings of the IEEE 2002, vol., no., pp.513,516, 2002.

[11] Mingcui Zhou; Wentai Liu; Sivaprakasam, M., "A closed-form delay formula for on-chip RLC interconnects in Current-mode signaling," Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on, vol., no., pp.1082, 1085 Vol. 2, 23-26 May 2005.

[12] Jaday, S.; Khanna, G.; Kumar, A.; Saini, G., "Low power high throughput Current-mode signalling technique for global VLSI interconnect," Computer and Communication Technology (ICCCT), 2010 International Conference on , vol., no., pp.290,295, 17-19 Sept. 2010.

[13] R. Bashirullah, W. Liu, and R. Cavin, "Current-mode signaling in deep submicrometer global interconnects,"IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 11, no. 3, pp. 406–417, Jun. 2003.

[14] International Technology Roadmap for Semiconductors (ITRS), 2009 update. [Online document], Available http: //public.itrs.net.

IOT Based Light Intensity Monitoring System using Embedded Linux & Raspberry Pi

KESANAPALLI ASKA RANI1 ,BODDU MANJULA

1.Assistant Professor, Dept of ECE, MRCE, HYDERABAD,Email:aksaranik@gmail.com.

2.Assistant Professor, Dept of ECE, MRCE,HYDERABAD , E-mail: manjulakoteswararao007@gmail.com.

Abstract: Accurate and quantifiable measurement of light is essential in creating desired outcomes in practical day to day applications as well as unique applications such as Traffic lighting system, Poultry Industry, Gardening, Museum lighting system, at emergency exits etc. Hence, Light measurement and analysis is an important step in ensuring efficiency and safety. Many of the industries are burdened with limited number of resources and real shortage of experts on their fields; real time remote monitoring presents an effective solution that minimizes their efforts and expenditures to achieve the desired results within time. This paper introduces real time remote Light intensity monitoring system using Raspberry Pi which enables the user to track the lighting system remotely. Raspberry pi is a low cost ARM11 processor Linux based computer which acts as a server, and it communicates with clients with LAN or external Wi-Fi module. The key feature of this system is light intensity being monitored instantaneously and data stored in the database for future use, and shown in the form of dynamic charts to the user according to the user requirement in a terminal device like Tablet or Smart Phone or any internet enabled device. This empowers experts to make right decisions at right time to get desired results.

Keywords: LDR. Light Intensity, Temperature, Smoke, Raspberry Pi, Web Server, Camera, Buzzer.

I.INTRODUCTION

There are many applications available to measuring and maintain the sufficient light levels such as laboratories, hospitals, educational institute, etc. To sustain healthier and safety environment enough light levels in the premises are needed. Without any distraction of whether condition, the light intensity has to be adequate for light levels intensity Some of important locations and light intensity is shown in Table I.

TABLE I. Optimum Average Light Intensity at Various Locations. Consider following Applications as an Example

Location	Illuminance(LUX)
Homes, Theaters	150
Library(Reading Area)	200
General Office work	500
Class room	300

A.Traffic Lighting System:- To ensure safety on the road, traffic lights need to be clearly visible for road users. The light intensity has to be sufficient under every (weather) condition, which set in legal standards. Over the course of time, the luminous intensity of traffic lights slowly decreases. Possible reasons are pollution of lenses or reflectors, aging of the light source

or individual LED failure. Remote monitoring enables the road authority to carry out timely services, in such a way that traffic lights keep satisfying the statutory rules for optimal traffic safety.

B. Poultry Industry:-Light Intensity is an important management factor in poultry industry to obtain optimal production. The intensity depends upon the age and type of housing being used, and type of chicken, be it broiler, breeder or layer. With blackout housing both male and female can be exposed to 3.5 fc from day one to day six and then placed on 1 fc to 19 or 20 weeks. After 19 - 20 weeks the broiler breeders can be exposed to about 3.0 to 5.0 fc during the entire production period. Layers should be exposed to about .5 to 1.5 fc (One foot-candle = 10.76 lux) for better production [4-6].

C. Plants Growth:-Deficient light intensity tend to reduce plant growth, development and yield. This is because low amount of solar energy restricts the rate of photosynthesis. Below a minimum intensity, the plant falls below the compensation point. Compensation point is the metabolic point at which the rates of photosynthesis and respiration are equal so that leaves do not gain or lose dry matter. Photosynthesis significantly slows down or ceases while respiration continues. Likewise, excessive light intensity should be avoided.

D. Museum Lighting System:- Light intensity is a primary consideration in museums to protect historic artifacts from damage. 5 to 10 foot-candles (approx. 50 to 100 lux) is currently considered to be the maximum allowable light level for very sensitive materials, such as prints, drawings,watercolors, dyed fabrics, manuscripts, and botanical specimens. Up to 15 footcandles.

E. Open CV: OpenCV is an open source library for image and video analysis, originally introduced more than decade ago by Intel. Since then, a number of programmers have contributed to the most recent library developments. The latest major change took place in 2009 (OpenCV 2) which includes main changes to the C++ interface. Nowadays the library has >;2500 optimized algorithms. It is extensively used around the world, having >;2.5M downloads and >;40K people in the user group. Regardless of whether one is a novice C++ programmer or a professional software developer, unaware of OpenCV, the main library content should be interesting for the graduate students and researchers in image processing and computer vision areas. To master every library element it is necessary to consult many books available on the topic of OpenCV. However, reading such more comprehensive material should be easier after comprehending some basics about OpenCV from this paper.

II. SYSTEM ARCHITECTURE

The system architecture of this proposed system is following.

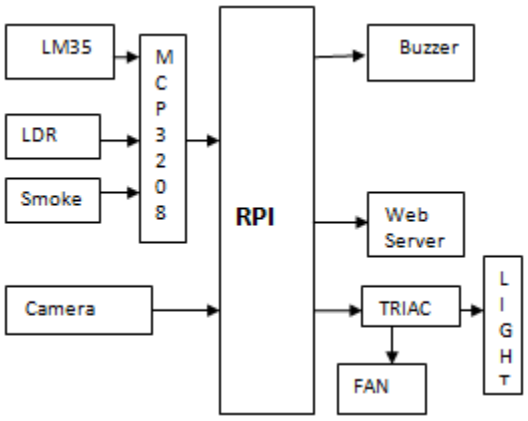


Figure.1. Block Diagram

Raspberry Pi: Hardware implementation for This proposed system is shown in above with the blocks. Raspberry Pi is the processor and its relevant components. The Wi-Fi is used for wireless communication and Wi-Fi USB module is interfaced to Raspberry Pi's USB port and sensor's data is to upload to web server and live monitoring by camera and Buzzer is used for alarm and LCD is used for display the Sensors data and TRIAC is used for switching the fan

and light. When the sensors are data reached threshold limit the buzzer sound will alert.

III. IMPLEMENTATION

A. Hardware

In hardware implementation, Raspberry Pi plays a key role in monitoring in this system. The Raspberry Pi is a small computer, same as the computers with which you're already familiar. It uses a many different kinds of processors, so can't install Microsoft Windows on it. But can install several versions of the Linux operating system that appear and feel very much like Windows. Simple to use but powerful, affordable and in addition difficult to break, Raspberry Pi is the perfect device for aspiring computer scientists. This small computer features amazing HD (high-definition) quality, video playback, also sports high quality audio and has the capability to play 3D games. The device use the ARM processor which does nearly all of the hard work in order to run the Raspberry Pi. The overview of Raspberry Pi has shown below

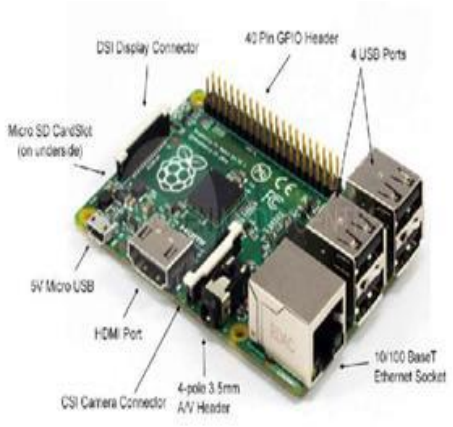


Fig.2. Raspberry Pi

1. GPIO: One powerful feature of the Raspberry Pi is the row of GPIO (general purpose input/output) pins along the Top edge of the board. These pins are physical interface between the pi and the oust side world. At the simplest Level ,you can think of them as switches that you can turn on or off(input) or that the pi can turn on or off(output).Of the 40 pins,26 are GPIO pins and other are power and ground pins. You can program the pins to interact in amazing ways with the real world. Inputs don't have to come from a physical switch. It could be input from a sensor or a signal from another computer or device, for example. The output can also do anything, from turning on LED to sending a signal or data to another device. If the Raspberry Pi is on a network, you can control devices that are attached to it from any where and those devices can send data back. Connectivity and control of physical devices over the internet is a powerful and exciting thing, and Raspberry Pi is ideal for this.

2. Temperature Sensor*:* The temperature sensor will give a variable output voltage with respect to the temperature variation. LM-35 is used as temperature sensor which is a precision integrated-circuit temperature sensor, Calibrated directly in ° Celsius (Centigrade), Linear + 10.0 mV/oC scale factor with accuracy O.soC (at +25°C) with rated for full -55° to +150°C range. The Temperature Sensor which I have used in this project has shown below:

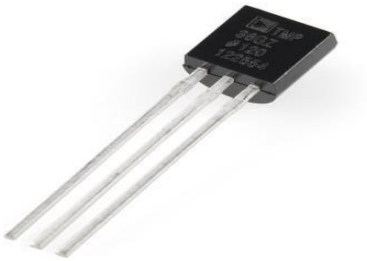


Fig.3. Temperature Sensor

3.Smoke Sensor*:* The smoke sensor will give a variable output voltage with respect to the temperature variation. There is better sensitivity for natural gas and coal gas. The Smoke Sensor which I have used in this project has shown below:

Fig.4. Smoke Sensor

4. LDR: LDR (Light Dependent Resistor) is variable resistor, the resistance of the LDR is inversely proportional to the light intensity, it exhibits maximum resistance in the absence of light and minimum resistance in the presence of light. The LDR which I've used in this project has shown below

Fig.5. LDR

5. MCP3208: MCP3208 devices are successive approximation 12-bit Analog-to-Digital (A/D) Converters with on-board sample and hold circuitry. The MCP3208 is programmable to provide two pseudo-differential input pairs or four single ended inputs. The MCP3208 is programmable to provide four pseudo-differential input pairs or eight single ended inputs. The ADC which I have used in this project has shown below: Differential Nonlinearity (DNL) is specified at ±1 LSB, while Integral Nonlinearity (INL) is offered in ±1 LSB (MCP3208-B) and ±2 LSB (MCP3204/3208-C) versions. Communication with the devices is accomplished using a simple serial interface compatible with the SPI protocol. The devices are capable of conversion rates of up to 100 kbps. The MCP3208 devicesoperate over a broad voltage range (2.7V - 5.5V).

B. CAMERA

.Fig.6. USB Camera

The USB camera Module is interfaced to the Raspberry Pi's USB port. The camera is mainly used to captured the changes in the environment i.e. Motions. The required power supply to operate USB camera will get it from Raspberry Pi only.

1. BUZZER: A buzzer or beeper is an audio signaling device which may be mechanical, electro mechanical, or piezoelectric. Typical uses of buzzers and beepers include devices. The buzzer which have used in this project is shown below fig:

Fig.7. Buzzer

2. TRIAC: The BT136 can be used in circuits of frequency conversion, voltage adjust and control. TRIAC's are widely used in AC power control applications. They are able to switch high voltages and high levels of current, and over both parts of an AC waveform. This makes triac circuits ideal for use in a variety of applications where power switching is needed. One particular use of TRIAC circuits is in light dimmers for domestic lighting, and they are also used in many other power control situations including motor control. The TRAIC which I have used in this project is shown below:

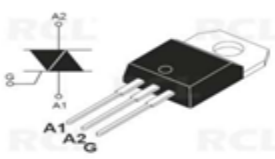


Fig.8. TRIAC.

The TRIAC is a development of the thyristor. While the thyristor can only control current over one half of the cycle, the TRIAC controls it over two halves of an AC waveform. As such the TRIAC can be considered as a pair of parallel but opposite thyristors with the two gates connected together and the anode of one device connected to the cathode of the other, etc. However the names of these are a little more difficult to assign, because the main current carrying terminals are connected to what is effectively a cathode of one thyristor, and the anode of another within the overall device. There is a gate which acts as a trigger to turn the device on. In addition to this the other terminals are both called Anodes, or Main Terminals These are usually designated Anode 1 and Anode 2 or Main Terminal 1 and Main Terminal 2 (MT1 and MT2). When using TRIAC's it is both MT1 and MT2 have very similar properties.

C. Software Here, to program Raspberry Python was used. And a Sever as HTML Web server. Final Schematic Diagram of this Project has shown below:

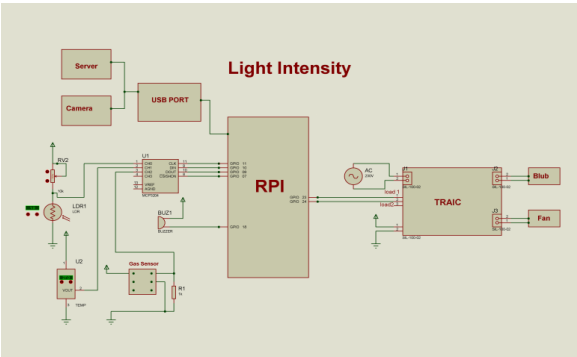


Fig.9. Schematic

IV. ALGORITHM & FLOWCHART

A. Algorithm

- Step– 1: Initialize RPI and camera.
- Step– 2: Taking sensor reading by ADC which have interfaced with RPI .
- Step – 3: uploaded the Sensor's data in to web server and Live streaming by camera which have interfaced with RPI .
- Step– 4:if the sensor's reached the threshold limits the light intensity will vary and fan will turn on and buzzer will turn on.
- Step– 5: continues till system runs.

B. Flowchart

The flowchart of this paper is shown below.

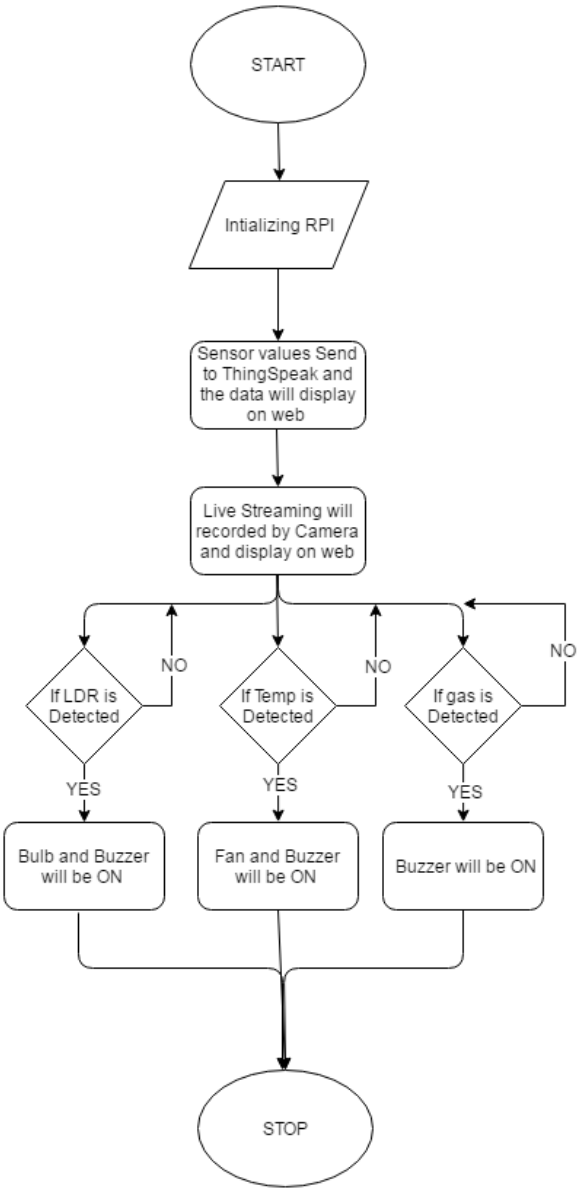


Fig.10. Flow Chart

V. RESULTS

Fig.11. Final Prototype

VI. CONCLUSION

The Facility manger will have skill, training and experience but lagging with lack of information to take action immediately. In the paper, we have proposed and developed cloud based light intensity, temperature and smoke monitoring system. This helps to Facility manger to take necessary action at right time, with proper controlling with can achieve desired results and we can monitor live streaming by camera. To evaluate the system, we have considered laboratory as an example but it can be used at various applications like traffic light monitoring, poultry lighting and museum lighting etc to avoid damages.

VII.REFERENCES

[1]http://www.engineersgarage.com/armprojects /introduction-to-arm-microcontroller-RaspberryPi [2]B.Johnson, “How the Raspberrypi works” Internet: computer.howstuffworks.com/raspberry-pi1.htm
[3]http://www.nex-robotics.com/RaspberryPi development-board/RaspberryPi-development-board.html
[4]https://en.wikipedia.org/wiki/Temperature-sensor [5]https://en.wikipedia.org/wiki/soil - sensor
[6]https://en.wikipedia.org/wiki/Rain fall-sensor

Smart Waste Collection Monitoring and Alert System via IoT

International Conference on Latest Trends in Electronics and Communication (ICLTEC)-ISBN 978-93-88808-62-0

Vaddepally Anusha
M.Tech Scholar
Department of Electronics and Communication Engineering
Mallareddy College of Engineering, Hyderabad, India
anushavaddepally18@gmail.com

***Abstract*—The uncollected waste material when the waste bin is full is a common problem nowadays. Thus, an efficient waste management for the waste material is essential in ensuring a clean and green surrounding environment. This paper presents an Internet of Things (IoT) based Smart Waste Collection Monitoring and Alert System to monitor the waste material at the selected site of garbage collection area. The system is implemented using an ultrasonic sensor which is connected to Arduino UNO as to monitor waste bin garbage level. In this system, waste bin depth level will be sent via Arduino Ethernet Shield with an Internet connection to the Ubidots IoT Cloud. The Ubidots store the collected waste bin level data into IoT database and display the waste bin depth level on online dashboard for real-time visualization. The Ubidots Event manager invoke a notification alert to garbage collector mobile phone via a SMS when the waste bin is nearly filled for immediate waste collection. Therefore, the waste collection became more effective and systematic.**

Keywords— Ultrasonic Sensor, Arduino UNO, Smart Waste Collection Monitoring and Alert System, Ubidots.

I. INTRODUCTION

Currently, over 23,000 tonnes of waste is produced each day in Malaysia. However, this amount is expected to rise to 30,000 tonnes by the year 2020. The amount of waste generated continues to increase due to the increasing population and development, and only less than 5% of the waste is being recycled. Despite the massive amount and complexity of waste produced, the standards of waste management in Malaysia are still poor. Hence, the country will face a problem when there is no proper management for this waste collection. As Malaysia is a developing nation and furthermore has comparative issues, for example, legitimate innovations, labour, arrive shortage and different offices which are deficient to adapt to the regularly expanding rate of waste age[1].

An efficient waste material collection is essential to prevent the waste from affecting human health and polluting environment. Large quantities of uncollected waste material are one of the thing that can affect society health and ecological system if it is not properly managed. The country will face a problem when there is no proper management for this waste collection department. As the country grow, waste material also grows too. Hence, one of the problem to solve the uncollected waste problem by knowing when the waste bin is full and ready for waste collection. This information can help the city council to properly schedule their garbage truck for waste collection within their managed area. This will improve the movement of waste collection fleet resources while enhance the efficiency of waste collection system.

Solid waste management is one of major aspect which has to be considered in terms of making urban area environment healthier. It is become necessary and challenging to manage the solid waste with rapid urbanization and increased population growth. An environment will be polluted and dirty if the waste material is not been manage and collected in time. A better waste management solution can helps improving the general wellbeing of a community and built up a better neighborhood. Nowadays, numerous IoT based solution for waste management are implemented to improve the collection of garbage which would ensure healthy environment for life on this green planet, with greater efficiency [2],[3],[4],[5]. Municipalities wanting to achieve cleaner urban environments can implement an IoT based solution [6],[7]. Some of the IoT based solutions [8],[9],[10] for waste management provide a notification alert when the garbage bin reaches its full capacity for immediate waste collection. An IoT based cost-effective system that can monitor the everyday garbage IoT based solid waste management system which enables garbage bin monitoring, dynamic scheduling and routing of garbage collector trucks in a smart city [11],[12]. A review of existing IoT-enabled solutions in smart cites' waste management is done here to bring together the state-of-the-art solution for example in term of self-powered solution.[13],[14].

IoT which is a new platform that is very useful for people in this world. IoT is the core of such revolutionary of growing engines. IoT is possible due to sufficient power supply and internet connectivity[15]. The term of IoT is commonly used to describe a framework where sensors are connected to objects and help these objects to share their ‘digital voice’ with the external world over internet connection. In the recent time, IoT has become a compilation of purpose-built networks. There were many IoT platforms such as Blynk, Ubidots, IBM Bluemix and Devicepilots as bidots is chosen as a platform of IoT for this project. Ubidots is a cloud service that offers a friendly and intuitive interface where the users can interact with a variety of devices, ranging from a cell phone or a computer, to an embedded system such as a microcontroller system. In a nutshell, Ubidots is a platform that allows to link different types of devices to a cloud database and save variables that can represent them in a simple and fast way and secure manner.

This paper presents a Smart Waste Collection Monitoring and Alert System (SWCMAS) using IoT technology on Ubidots Cloud. The remainder of the paper is organized as follows. Section II describes the system development and overall design approach. Section III discussed the experimental result of an IoT smart waste collection monitoring and alert system performance. Finally, section IV provides the concluding remarks and point out the ideas for future extension of this work.

II. System Development

In this section, a brief explanation regarding on the project development and methodology will be described. This project proposed a system to control a waste material from overflow from the waste bin and alert is send to cleaner for waste collection. Using the anticipated system, monitoring of the waste collection status could be monitored effectively. This project designates a technique in which could monitor the garbage level at regular intervals as overflow of the bin can be prevented. The filling level of the garbage in the dustbin and its original level height could be sensed/ monitored by the ultrasonic sensor. Programming in the Arduino UNO is done in such a way that once a particular level of filling is sensed, information as a message is sent to the user, requesting for cleaning of the dust bin.

Referring to the block diagram in Figure 1 below, the developed system consists of a 1) sensor node that implemented using an Arduino Uno board connected with Arduino Ethernet Shield, HCRS-04 ultrasonic sensor and buzzer; 2) Wired router that interconnected the sensor node to Ubidots IoT Cloud platform and 3) Ubidots Cloud platform that consist secured IoT devices organization.

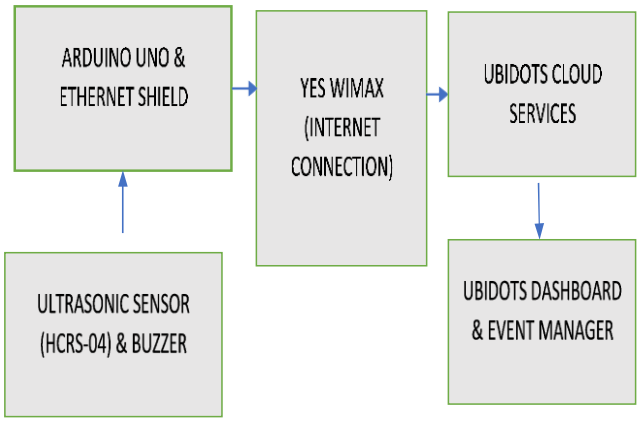


Fig. 1. Block diagram of SWCMAS

Ubidots Dashboard to display and visualize the waste bin depth level data and Ubidots Event manager to notify an alert to consumer via telegram/sms on smartphone when the waste bin is nearly filled.

The system development consists of hardware and software development phase. In hardware development phase, the Arduino Uno is use as a microcontroller for the sensor node. The sensor node is interface with Ultrasonic sensor for detection of the waste bin depth.

A. System Hardware Development

Figure 2 above shows the schematic diagram for connection of Arduino Uno and Ultrasonic sensor. The pin that is used is all digital pins, hence no usage of analog pins. The ground pin of ultrasonic sensor must be connected to ground pin of Arduino Uno to avoid short-circuit. The trig pin and echo pin is connected to (digital) pin 8 and (digital) pin 9 respectively. The RJ45 cable also is connected from Arduino Ethernet Shied to YES WiMAX router for wired internet connection.

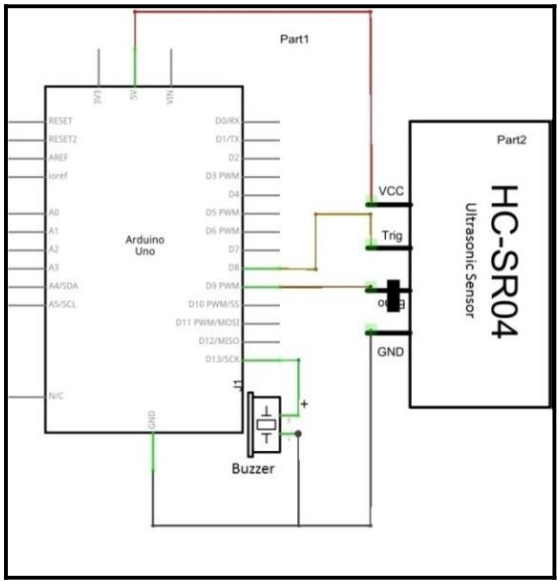


Fig. 2. Schematic diagram of SWCMAS

The ultrasonic sensors are placed at the top of the bin or at the lid of the waste bin as shown in Figure 3. When the waste is started to be filled, the Ultrasonic sensor starts emitting sound waves. At one side is the transmitter and other is the receiver, which measures distance by sending out a sound wave at a specific frequency and listening for that sound wave to bounce back. By recording the elapsed time between the sound wave being generated and the sound wave bouncing back, it is possible to calculate the distance between the ultrasonic sensor and the object. It comes as a whole with an ultrasonic transmitter and collector module. The distance can be calculated with the following formula:

$$\text{Distance } L = 1/2 \times T \times C \quad (1)$$

where L is the distance, T is the time between the emission and reception, and C is the sonic speed. (The value is multiplied by 1/2 because T is the time for go-and-return distance.).

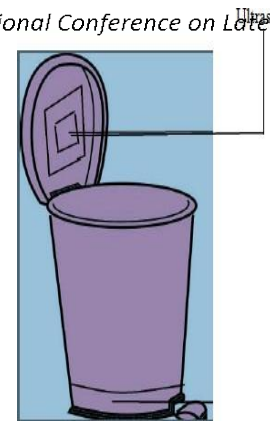


Fig. 3. Waste Bin Model

Figure 4 shown the waste bin at the selected site. The ultrasonic sensors are placed at the top of the bin. The step of the collection data will be repeated as using the waste bin model. The data or information is collected and send to Ubidots platform.

Fig. 4. Waste Bin at Selected Site

B. Software Development and Configuration

In software development and configuration phase, the MAC and IP address of sensor node have to be obtained and configure into Arduino program for sending the waste bin depth level to Ubidots IoT Cloud.

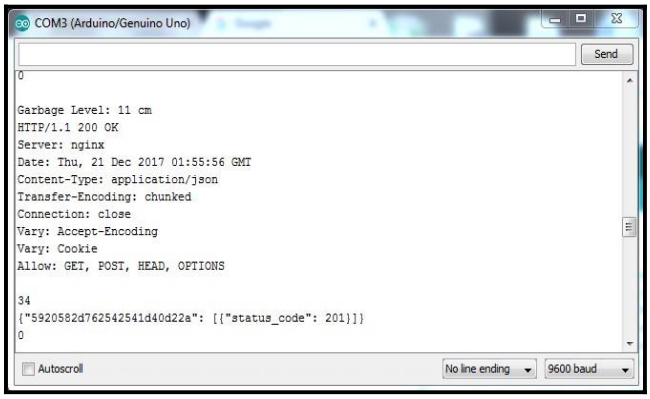


Fig. 5. Ultrasonic sensor data publish to Ubidots Cloud

Figure 5 shows the the waste bin depth level data collection from the sensor node and publish the collected data using MQTT protocol connection to Ubidots platform in Arduino IDE. Figure 5 also shows that the transfer-encoding for this project is chunked and vary in accept-encoding and cookie. The connection is established as the connection from the Arduino IDE and Ubidots website is close.

After many steps of microcontroller coding using Arduino, Node configuration, Arduino Ethernet Shield configuration, the data is finally collected and sent to Ubidot IoT Cloud. Then, the project proceeded by configuring Ubidots platform IoT organization. The interface of Ubidots is user-friendly and ensures quick visualization of the multiples of interest.

By using this platform, you can create an event or warning messages, send them to a mobile device and configure the device to execute a control action. The system will alert the user and will send a message to user through cell phone via Wi-Fi. The device must be built in the Ubidots platform to take the devices API ID as to synchronize with Arduino IDE coding. Within Ubidots, the data can be retrieved in an instant. The event will be created as shown as Figure 6 below to notify the user when the waste value is less than 4 cm.

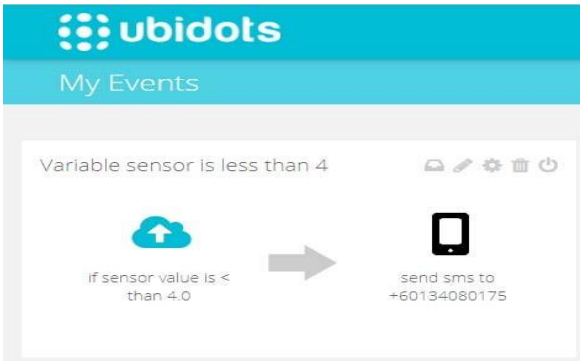


Fig. 6. The Event Setup for SMS notification Alert

III. RESULT

This section shown the results of data gathered during this project implementation. All of the data and information details were collected to evaluate the system performance. The result consists of two parts. The first part is an Ultrasonic sensor distance result that indicate the depth level of waste bin data display in the serial monitor on Arduino IDE. The second part is obtained from depth level of waste bin data collection through IoT Cloud on Ubidots Dashboard which is an online display. By creating a device in Dashboard, we can collect and store the waste bin depth level data on the specified device. The notification alert can be invoked when certain value of device data changes more or less from the specified value.

A. Arduino IDE results

Figure 7 shows the result on the serial monitor on Arduino IDE that indicate the depth level of waste bin data.

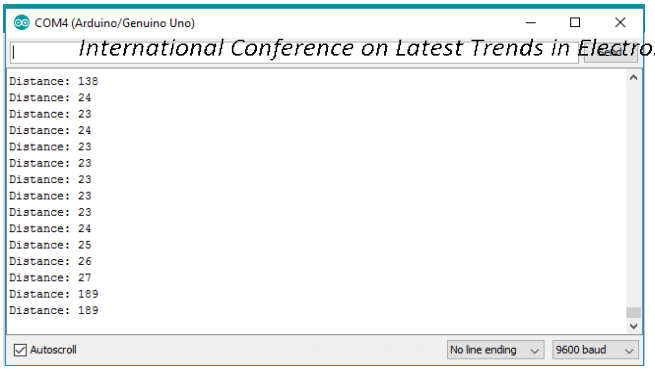


Fig. 7. The Ultrasonic Sensor Distance Measurement Result

B. *Ubidots Dashboard results*

The performances of this waste collection monitoring system can be monitored through Ubidots Dashboard. The data or information of the distance value can be view in raw, average or overall. The data and information that is collected is from waste bin from the selected site. The data will be sends from Sensor node through Arduino Ethernet Shield to Ubidots devices that had been created. Figure 8 shows the graph of the collected data of distance between garbage level and the waste bin lid while Figure 9 shows the value of the depth of the garbage level to the lid chronological order as the waste bin is nearly full.

Fig. 8. Waste bin depth level data graph at Ubidots

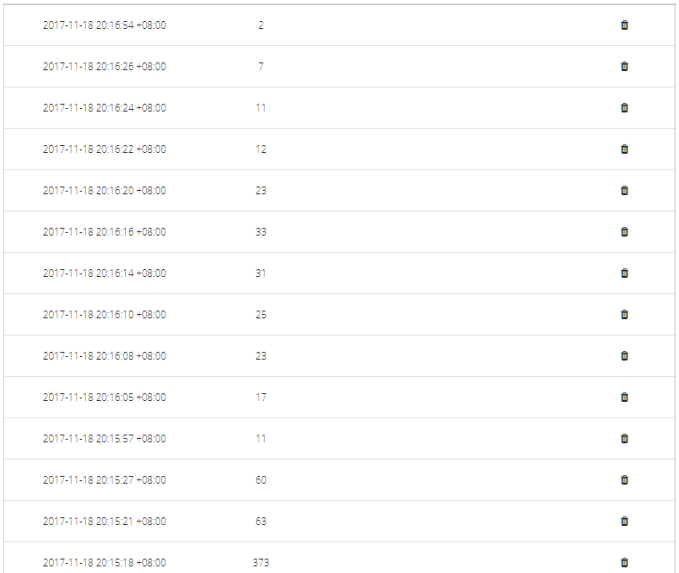
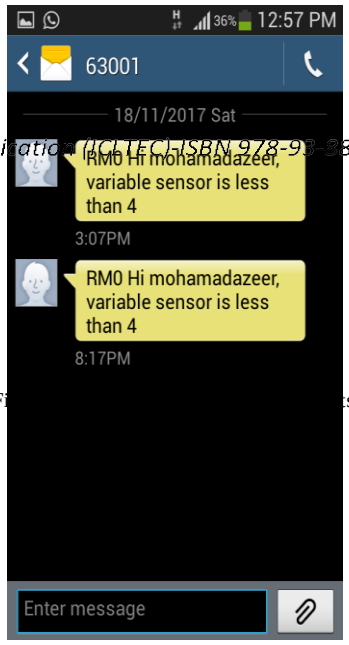


Fig. 9. The Waste bin depth level on Ubidots Cloud



Fi ts.

As the garbage level increase, the distance between the waste bin lid with ultrasonic sensor will decrease Once the level of distance less than the determined value (4 cm), Ubidots will alert and send message to authorized person in form of SMS as shown in Figure 10.

IV. Conclusion

In this paper, we propose a new solution to enhance waste collection efficiently using the Arduino Uno with Arduino Ethernet Shield technology and ultrasonic sensor systems. In this proposed system, the garbage overflow of garbage can be avoided and managed efficiently. This will intimate or send SMS or email to the authorized person through Ubidots platform. The garbage managing system and the facility of collecting the garbage presently doesn't fit to the current requirement. Hence better facility of collecting garbage and transportation should be provided. Since, this system provides the information when the bin gets completely filled with garbage, it reduces the number of times the arrival of vehicle which collects the garbage. This method finally helps in keeping the environment clean. Thus, the waste collection is made more efficient.

Acknowledgment

We would like to thanks the sponsors of development of Ubidots for free evaluation license on Industrial IoT and also provide us with the reading materials supported hardware libraries.

REFERENCES

[1] J. Sreenivasan, M. Govindan, M. Chinnasami, and I. Kadiresu, "Solid Waste Management in Malaysia – A Move Towards Sustainability," *Waste Manag. An Integr. Visions*, vol. 2005, no. April 2005, pp. 55–70, 2012.

[2] M. Adam, M. E. Okasha, O. M. Tawfeeq, M. A. Margan and B. Nasreldeen, "Waste Management System Using IoT," 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, 2018, pp. 1-4

[3] S. Aleyadeh and A. M. Taha, "An IoT-Based Architecture for Waste Management," 2018 IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, 2018, pp. 1-4.

[4] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[5] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[6] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[7] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[8] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[9] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[10] S. K. Memon, F. Karim Shaikh, N. A. Mahoto and A. Aziz Memon, "IoT based smart garbage monitoring & collection system using

[11] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[12] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[13] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[14] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[15] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[16] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[17] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[18] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[19] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[20] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[21] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[22] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[23] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[24] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[25] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[26] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[27] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[28] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[29] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[30] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[31] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[32] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[33] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[34] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things (IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[35] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[36] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[37] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[38] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[39] G. K. Shyam, S. S. Manvi and P. Bharti, "Smart waste management using Internet-of-Things(IoT)," 2017 2nd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2017, pp. 199-203.

[40] W. Chen, Y. Wang, P. Huang, Y. Huang and M. Tsai, "A Smart IoT System for Waste Management," *2018 1st International Cognitive Cities Conference (IC3)*, Okinawa, 2018, pp. 202-203.

[41] S. S. Chaudhari and V. Y. Bhole, "Solid Waste Collection as a Service using IoT-Solution for Smart Cities," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-5.

[42] E. Al-Masri, I. Diabate, R. Jain, M. H. Lam and S. Reddy Nathala, "Recycle.io: An IoT-Enabled Framework for Urban Waste Management," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5285-5287

[43] S. Paul, S. Banerjee and S. Biswas, "Smart Garbage Monitoring Using IoT," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 1181-1185.

[44] H. N. Saha, S. Gon, A. Nayak, S. kundu and S. Moitra, "Iot Based Garbage Monitoring and Clearance Alert System," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2018, pp. 204-208.

[45] S. K. Memon, F. Karim Shaikh, N. A. Mahoto and A. Aziz Memon, "IoT based smart garbage monitoring & collection system using WeMos & Ultrasonic sensors," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-6.

[46] R. N. Subrahmanian, S. K. S, T. Suvidharbabu, and B. Manikandan, "Smart Bin using IoT with Fog Computing," International Journal of Emerging Technology in Computer Science & Electronics vol. 24, no. 6, pp. 13–16, 2017.

[47] S. Murugaanandam, V. Ganapathy and R. Balaji, "Efficient IOT Based Smart Bin for Clean Environment," 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 0715-0720.

[48] N. Alsbou, M. A. Samad, M. Alhashem and A. S. A. Abuabed, "Developing a Self-Powered Enlarging Smart Waste Bin," 2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC), Limassol, 2018, pp. 683-689.

[49] S. Mdukaza, B. Isong, N. Dladlu and A. M. Abu-Mahfouz, "Analysis of IoT-Enabled Solutions in Smart Waste Management," IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, 2018, pp. 4639-4644.

[50] D. Evans, "The Internet of Things How the Next Evolution of the Internet" Cisco Internet Business Solutions Group (IBSG) no. April, 2011. http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf

Chip Design for Turbo Encoder Module for In-Vehicle System

PULUKURI NEIL SAMUEL SADANAND,MTECH(ES&VLSI),MRCE

***Abstract*—This paper studies design and implementation of the Turbo encoder to be an embedded module in the in-vehicle system (IVS) chip. Field programmable gate array (FPGA) is employed to develop the Turbo encoder module. Both serial and parallel computations for the encoding technique are studied. The two design methods are presented and analyzed. Developing the parallel computation method, it is shown that both chip size and processing time are improved. The logic utilization is enhanced by 73% and the processing time is reduced by 58%. The Turbo encoder module is designed, simulated, and synthesized using Xilinx tools. Xilinx Zynq-7000 is employed as an FPGA device to implement the developed module. The Turbo encoder module is designed to be a part of the IVS chip on a single programmable device.**

***Index Term:* Turbo encoder module; field programmable gate array; emergency call; in-vehicle system chip.**

I. Introduction

The European emergency call (eCall) system is a telematics system designed to save more lives in vehicle accidents. It is a governmental mandatory system that is to be implemented by March 2018 [1][2]. The EU eCall system provides an immediate voice and data channel between the vehicles and an emergency center after car accidents. The data channel provides the emergency center with the necessary data for emergency aids.

The EU eCall system main parts includes the in-vehicle system (IVS), the public safety answering point (PSAP), a cellular communication channel. The IVS activates the data channel automatically when a car accident occurs. The IVS collects the minimum set of data (MSD) that includes GPS coordinates, the VIN number, and all required data for an emergency aid. It sends the MSD to the closest PSAP through a cellular channel in up to 4 seconds [1]. The PSAP sends the emergency team to the location of the accidents.

The IVS modem employs multiple modules for the MSD signal processing. The modules of the IVS are shown in Figure 1. The IVS employs a Turbo encoder as a forward error correcting (FEC) [1]. The Turbo encoder implements the digital data encoding technique in data transmissions. Turbo coding is one of the most popular and efficient coding technique to improve bit error rate (BER) in digital communications [3] [4]. The cyclic redundancy check (CRC) [5], the modulator [6], the demodulator-decoder [7] modules are projected and implemented on an FPGA device. They are developed to be embedded modules of the IVS chip.

978-1-5386-4649-6/18/$31.00

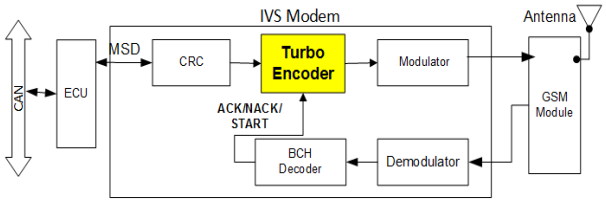


Fig. 1: The IVS block diagram.

This work studies the hardware development of the Turbo encoder. It employs FPGA technologies to develop the Turbo encoder to be an embedded module in the IVS modem. It discusses serial and parallel computation techniques for the Turbo encoder. It does not only design and implement the Turbo encoder module, but also proposes a better solution for the turbo encoder implementation. The improvement of the chip size and processing time are exhibited by developing the parallel computation technique for the Turbo encoder.

II. Turbo Encoder Module

The turbo encoder technique is one of the most powerful FEC techniques in digital communication [8]. The IVS employs a Turbo encoder module with $1/3$ code rate. The Turbo encoder functionalities are detailed in the third generation partnership project (3GPP) standards. The 3GPP Turbo encoder is illustrated in Figure 2 [8]. The input signal of the turbo encodes is the MSD data appended with the CRC parity bits in binary. The block length of the MSD data is 1148 bits. The output of the module is the MSD encoded data in binary. Implementing the turbo coding technique with $1/3$ coding rate and thrills bits, the length of the output is 3456 bits. The thrills structure has an impact of the Turbo encoder [9].

The Turbo encoder employs a parallel concatenated convolutional code (PCCC). The PCCC uses two constituent encoders with eight states as it is shown in Figure 2. The initial status of the register are zeros. The first constituent takes the MSD bits and implements the employed convolutional technique. It takes one bit at a time and generates one bit of

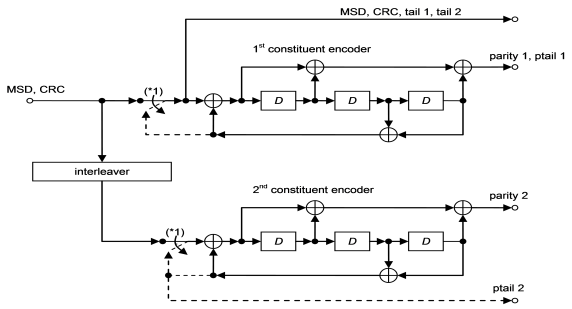


Fig. 2: The structure of the Turbo encoder.

parity1 bits. The second constituent implements an identical technique of the first constituent, but it calls for the MSD bit after they are interleaved with a 3GPP designed interleaver technique [8].

The length of the input data, parity1, and parity2 are 1148 bits. There are 12 bits of the tail bits. They are driven from the shift register feedback. The tail bits are applied for end points between the encoded data blocks. The output structure of the Turbo encoder is illustrated in Figure 3.

MSD+CRC	$tail_1$	$tail_2$	Parity 1	$ptail_1$	Parity 2	$ptail_2$

Fig. 3: The output buffer of the Turbo encoder.

Interleaver

Denote the transfer function of the employed PCCC as:

$$G(D) = \left(1, \frac{g_1(D)}{g_0(D)}\right) \tag{1}$$

where

$$g_(D) = 1 + D^2 + D^3$$

$$g_(D) = 1 + D + D^3$$

And denote the input bits to the encoder as $x_1, x_2, ..., x_K$, the output of the interleaver as $x'_1, x'_2, ..., x'_K$, and the output bits of the first and second constituents as $z_1, z_2, ..., z_K$ and $z'_1, z'_2, ..., z'_K$, respectively; where K is the number of input bits to the Turbo encoder.

The encoder output is expressed as:

$$d_K^{(0)} = x_K, d_K^{(1)} = z_K, d_K^{(2)} = z'_K$$

where $K = 0, 1, ..., K-1$.

The three code blocks of the output, $d_K^{(0)}$, $d_K^{(1)}$, and $d_K^{(2)}$ are separated by trellis bits. The trellis bits are generated from the tail bits of the shift registers after encoding of all the input bits. In figure 2, when the upper switch is lowered and the second constituent is disabled, the three tail bits are used to terminate the first constituent. The output bits of the Turbo encoder, including the trellis bits can be expressed as:

$$d_K^{(0)} = x_K, \ d_{K+1}^{(0)} = z_{K+1}, \ d_{K+2}^{(0)} = x'_K, \ d_{K+3}^{(0)} = z'_{K+1}$$

$$d_K^{(1)} = z_K, \ d_{K+1}^{(1)} = x_{K+2}, \ d_{K+2}^{(1)} = z'_K, \ d_{K+3}^{(1)} = x'_{K+2}$$

$$d_K^{(2)} = x_{K+1}, d_{K+1}^{(2)} = z_{K+2}, d_{K+2}^{(2)} = x'_{K+1}, d_{K+3}^{(2)} = z'_{K+2}$$

where $K = 0, 1, ..., K-1$.

The internal interleaver of the 3GPP Turbo encoder is designed to generate a systematic relationship between x_K and x'_K for any $40 \leq K \leq 5114$ [8]. There is a specific approach to design an internal interleaver for the employed Turbo encoder that is detailed in [8]. This work employs the 3GPP standard approach to design the internal interleaver for the employed Turbo encoder.

First, the input bits of the Turbo encoder is re-arranged in a matrix form that has column, C, and row, R. The rows are labeled as $0, 1, ..., R-1$ and the columns are organized as $0, 1, ..., C-1$. The numbers of rows and columns are determined according to the 3GPP standard for Turbo encoder interleavers [8].

Then the input bits $x_1, x_2, ..., x_K$ are re-organized in a matrix where $y_k = x_k$ for $k = 1, 2, ..., K$, and $y_k = 0$ for the elements that $R \times C > K$:

$$\begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_C \\ y_{(C+1)} & y_{(C+2)} & y_{(C+3)} & \cdots & y_{(C+C)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{((R-1)C+1)} & y_{((R-1)C+2)} & y_{((R-1)C+3)} & \cdots & y_{(R\times C)} \end{bmatrix}$$

Then an intra-row and inter-row permutation is performed on the $R \times C$ matrix. This work employs the 3GPP standard approaches for the intra-row and inter-row.

Denote,

$$s(j) = (v \times s(j-1)) \mod p \tag{2}$$

where $\left\langle s(j) \right\rangle$ for $j \in 1, 2, ..., p-2$ and $p(0) = 0$ is the inta-row permutation sequence and v is associated primitive root for the specified p from table **??**, and use table I to choose the appropriate pattern to compute the inter-row permutation, $\left\langle T(i) \right\rangle$ for $i \in 0, 1, ..., R-1$. i is the row number index of $R \times C$ matrix, and j is the column number index of the matrix.

Also the minimum prime integer (q_i) is determined in the sequence $\left\langle q(i) \right\rangle$ for $i \in 0, 1, ..., R-1$ such that $q_i > q_(i-1)$, $q_i > 6$ and g.c.d$(q_i, p-1) = 1$, where g.c.d is the greater common divisor.

Then one can build a sequence of the permuted prime integers $\left\langle r(i) \right\rangle$ for $i \in 0, 1, ..., R-1$ such that,

$$r_T(i) = q_i, i = 0, 1, ..., R-1$$

TABLE I: 3GPP inter-row permutation pattern

K	*R*	Inter-row permutation patterns <*T*(0), *T*(1), ..., *T*(*R* - 1)>
($40 \leq K \leq 159$)	5	<4, 3, 2, 1, 0>
($160 \leq K \leq 200$) or ($481 \leq K \leq 530$)	10	<9, 8, 7, 6, 5, 4, 3, 2, 1, 0>
($2281 \leq K \leq 2480$) or ($3161 \leq K \leq 3210$)	20	<19, 9, 14, 4, 0, 2, 5, 7, 12, 18, 16, 13, 17, 15, 3, 1, 6, 11, 8, 10>
K = any other value	20	<19, 9, 14, 4, 0, 2, 5, 7, 12, 18, 10, 8, 13, 17, 3, 1, 16, 6, 15, 11>

Denote the pattern of $i-th$ row Intra-row permutation as,

$\left\langle U_i(j) \right\rangle$ for $i \in 0, 1, ..., R-1$,
one can perform the intra-row permutation such that the position of the $i-th$ permuted bit of the $j-th$ row ($U_i(j)$) is calculated based on equation 3, 4, or 5.

if $(C = p)$ then,

$$U_i(j) = s((j \times r_i) \mod (p-1)) \tag{3}$$

where $j = 0, 1, ..., (p-1)$ and $U_i(p-1) = 0$.

if $(C = p+1)$ then,

$$U_i(j) = s((j \times r_i) \mod (p-1)) \tag{4}$$

where $j = 0, 1, ..., (p-1)$, $U_i(p-1) = 0$, and $U_i(p) = p$.

if $(C = p-1)$ then,

$$U_i(j) = s((j \times r_i) \mod (p-1)) - 1 \tag{5}$$

where $j = 0, 1, ..., (p-1)$.

And then the inter-row permutation is implemented on the $R \times C$ matrix by using the sequence pattern $\left\langle T(i) \right\rangle$ for $i \in 0, 1, ..., R-1$.

After the permutations, the elements of the $R \times C$ matrix is denoted by $y'_k = y_k$ such that:

$$\begin{bmatrix} y'_1 & y'_{(R+1)} & y'_{(2R+1)} & \cdots & y'_{((C-1)R+1)} \\ y'_2 & y'_{(R+1)} & y'_{(2R+2)} & \cdots & y'_{((C-1)R+2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y'_R & y'_{(2R)} & y'_{(3R)} & \cdots & y'_{(R \times C)} \end{bmatrix}$$

The output of the interleaver, $x'_2, x'_2, ..., x'_K$, is the bit sequence that are read out from the $R \times C$ matrix column by column starting with y'_1 and ending with $y'_{(R \times C)}$. The appended zero bits for $R \times C > K$ are removed from the output.

There are 1148 elements in the interleaver matrix. The 1148 bits of the MSD data are the elements of the interleaver matrix. The interleaver reorganizes the MSD bits in a systematic order. The intra-row and inter-row techniques of the interleaver elements organizations.

Denote the MSD bits as $B_1, B_2, ..., B_K$, where $K = 1148$. According to the algorithm that is explained in the above mathematical modeling, the interleaver matrix is a rectangular matrix, where R is the number of rows and C is the number of columns of the matrix. The interleaver matrix is designed for an input data block that consists of 1148 bits. As a result, the size of the matrix is 20 X 58. The following steps are implemented to drive the interleaver matrix:

1. The input bits of the matrix are denoted as $b_1, b_2, ..., b_K$, where $b_K = B_K$ and $K = 1148$. The remaining elements are padded with zeros.

2. The intra-row and inter-row permutation are performed according to 3GPP.

3. The calculated elements of the interleaver matrix, except for the padded bits, are stored in a file in hexadecimal format.

The Turbo encoder is developed in Verilog HDL language. Verilog has ability to read the hexadecimal file to get the data and use it as the interleaver data.

The length of the input data, parity1, and parity2 are 1148 bits. There are 12 bits of the tail bits. They are driven from the shift register feedback. The tail bits are applied for end points between the encoded data blocks. The output structure of the Turbo encoder is illustrated in Figure 3.

The employed interleaver for the Turbo encoder is designed according to 3GPP standards [8]. The interleaver elements are organized in a rectangular matrix. There are 1148 elements in the interleaver matrix. The 1148 bits of the MSD data are the elements of the interleaver matrix. The interleaver reorganizes the MSD bits in a systematic order. There are intra-row and inter-row techniques of the interleaver elements organizations.

Denote the MSD bits as $B_1, B_2, ..., B_K$, where $K = 1148$. According to the 3GPP standards [8], the interleaver matrix is a RxC rectangular matrix, where R is the number of rows and C is the number of columns. The size of the matrix is 20 X 58. The following steps are implemented to drive the interleaver matrix:

1. The input bits of the matrix are denoted as $b_1, b_2, ..., b_K$, where $b_K = B_K$ and $K = 1148$. The remaining elements are padded with zeros.

2. The intra-row and inter-row permutation is performed according to 3GPP [8].

3. The calculated elements of the interleaver matrix, except for the padded bits, are stored in in a file in hexadecimal format.

The Turbo encoder is developed in Verilog HDL language. Verilog has ability to read the hexadecimal file to get the data and use it as the interleaver data.

III. FPGA Design for the Turbo Encoder Module

FPGA technologies are employed to develop and implement the designed Turbo encoder module. The register transfer level (RTL) of the module is developed in Verilog HDL. There are multiple registers defined for the input, output, and necessary parameters to implement the Turbo encoding technique. This work studies two methods to execute the encoding, which are serial computation and parallel computation.

The serial computation method processes one bit in one clock cycle. It reads the input data of the MSD, builds the input and output registers, and calculates the parity1, parity2, and the tail bits in a serial process. After performing the encoding, it generates the output bits. Although the method is designed and implemented, it is noted that there is a long processing time that can be overlapped with the other processes in the module. Figure 4 shows the pseudocode of the serial computation of the Turbo encoder module.

The parallel computing technique is employed to develop the Turbo encoder in Verilog. There are many processes in the serial computation technique that are overlapped by using parallel computing technique. There are two functions developed in the parallel Turbo encoder. The two functions implements almost all the processing time of the encoding technique. The Turbo encoding technique needs the MSD data as a whole package to implement the encoding. Figure 5 shows the pseudocode of the implemented parallel technique

```
Pseudocode code for Serial Computation of Turbo encoder
module TURBO_SERIAL ( inputs, outputs;)
        define REGISERS and PARAMETERS;
      always @(posedge clock, posedge reset)
        begin
         if (reset) output=0;
         else begin
            repeat1 (1148) {
                      Read MSD input data}
             If (repeat1 is done)
                repeat2 (1148) {
                      Build output register for MSDinput;
                      Build output register for Parity1;
                      Build output register for Parity2; }
             If (repeat2 is done)
                repeat3 (1148) {
                      Process Parity1 bits; }
             If (repeat3 is done)
                Repeat4 (3) {
                      Process tail1 bits;
                      Process ptail1 bits;    }
             If (repeat4 is done)
                Repeat5 (1148) {
                      Process Parity2 bits; }
             If (repeat5 is done)
                Repeat6(3) {
                      Process tail2 bits;
                      Process ptail2 bits;    }
             If (repeat6 is done)
                Repeat7(3456) {
                      Generate output bits    }
              end  end
endmodule;
```

Fig. 4: The pseudocode for serial computation of the Turbo encoder.

for the Turbo encoder. Both Pseudocodes of the serial and parallel computing techniques are designed in Verilog and implemented on an FPGA device.

The processing time of the Turbo encoder module (in clock cycles) is denoted by T_s for the serial computation and by T_p for the parallel technique, one has,

$$T_s = T_r + T_b + T_{parity1} + T_{tail1} + T_{parity2} + T_{tail2} + T_w \quad (6)$$

$$T_s = 1148 + 1148 + 1148 + 3 + 1148 + 3 + 3456 = 8054$$

where T_r is the time for reading the 1148 bits of the MSD, T_b is the processing time to build the output register, T_{parit1} is the time of processing parity bits, T_{tail} is the time of processing tail bits, and T_w is the time of generating output bits.

Note that the $T_r + T_b + T_{parity1} + T_{tail1} + T_{parity2} + T_{tail2}$ are processed in one clock cycle in the parallel computing technique.

$$T_p = 1 + T_w = 1 + 3456 = 3457 \quad (7)$$

Then one has,

$$T_p = 0.42 T_s \quad (8)$$

Eq. 8 reveals that the parallel computing can improve the proceeding time of the Turbo encoder by 58%.

A. Simulation and Verification

Xilinx tools are utilized to simulate the developed modules. A test bench is designed to simulate the Turbo encoder. Verilog

```
Pseudocode code for Parallel Computation of Turbo encoder
module TURBO_PARALLEL ( inputs, outputs;)
        define REGISERS and PARAMETERS;
 function [3455:0]  codedMSD1;
              input [1147:0] MSDdata;
                 begin
                    repeat1 (1148) {
                       Build output register for MSDinput;
                       Build output register for Parity1;
                       Build output register for Parity2; }
                  If (repeat1 is done)
                     repeat2 (1148) {
                      call  CodedMSD2 (codedMSD1) }
                  end
       endfunction
 function [3455:0]  codedMSD2;
              input [3455:0] codedMSD1;
                 begin
                 repeat3 (1148) {
                       Process Parity1 bits; }
                  If (repeat3 is done)
                     repeat4 (3) {
                       Process tail1 bits;
                       Process ptail1 bits;    }
                 If (repeat4 is done)
                    repeat5 (1148) {
                       Process Parity2 bits; }
                 If (repeat5 is done)
                    repeat6(3) {
                       Process tail2 bits;
                       Process ptail2 bits;    }
                  end
       endfunction
   always @(posedge clock, posedge reset)
      begin
       if (reset) output=0;
       else begin
          repeat1 (1148) {
                    Read MSD input data    }
              If (repeat1 is done) {
                  call  CodedMSD1 (MSDdata)   }
              Repeat7 (3456) {
                    Generate output bits    }
            end  end
endmodule;
```

Fig. 5: The pseudocode for parallel computation of the Turbo encoder.

HDL is employed to design the test bench. There are two MSD data that are simulated for each of the serial computation and parallel computation of the Turbo encoder. Figure 6 shows the simulation results for the serial computation of the Turbo encoder module. The simulation indicates that the structure output data is correct. However, there is a long time that can be overlapped, which is colored in red in the output trace.

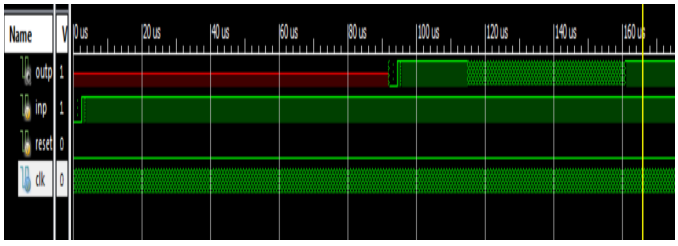


Fig. 6: The simulation result of the Turbo encoder with serial computation.

Figure 7 shows the simulation result of the parallel computing technique. The module starts to generate the output in the beginning. Note that the MSD is encoded in a shorter time compare with the serial computation simulation. This is the

result of parallel computing of multiple process in one clock cycle.

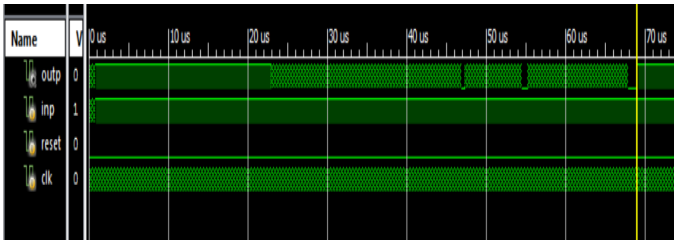


Fig. 7: The simulation result of the Turbo encoder with parallel computation.

B. Hardware Implementation

FPGA has been widely used in the SoC design [10][11]. Zynq-7000 FPGA device is employed to implement the developed Turbo encoder module. Both serial and parallel computation methods are implemented on the FPGA device separately. The Turbo encoder module is synthesized and loaded on the FPGA device. The synthesis reports show that the parallel computation does not only save processing time, but also reduces the utilized logics and the chip size. Figures 8 and 9 show the utilization flip flops and look-up tables (LTU) for the serial and parallel computation respectively.

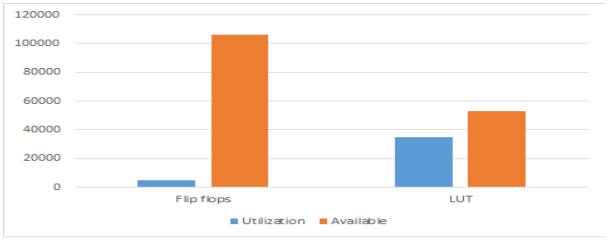


Fig. 8: The logic utilization of the Turbo encoder with serial computation.

The impact of the parallel computation on the logic utilization is shown in Figure 10. Note that the utilized flip flop and LUTs are significantly reduced when parallel computation is employed. The LUT utilization is improved by 73% and the flipflop utilization enhanced by 75%. It is proven that the FPGA technologies strongly supports parallel computation for the Turbo encoder. By reducing the size of the Turbo encoder module, the IVS can be implemented on a single programmable chip with less hardware constrains.

IV. Conclusion

The Turbo encoder module is designed and implemented to be an embedded module in the IVS modem. FPGA technologies are employed to develop the Turbo encoder module. Xilinx tools and Verilog HDL are employed to design and simulate the module. Both serial and parallel computation techniques are studied for the encoding process. It is shown that the parallel computation can improve the chip size and

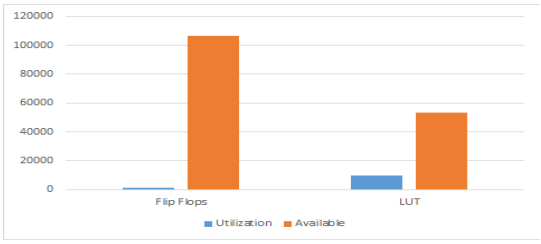


Fig. 9: The logic utilization of the Turbo encoder with parallel computation.

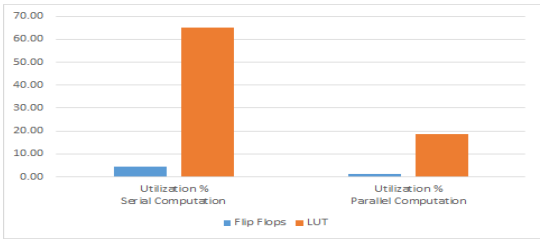


Fig. 10: The logic utilization of the Turbo encoder, serial vs parallel computation.

processing time of the module. Comparing with the serial computation technique, the parallel computation encoding, improves the processing time by 58% and logic utilization by 73%. The processing time enhancement can be seen in both simulation and analyzing the chip processing.

References

[1] "eCall data transfer; in-band modem solution; general description," 3GPP, Tech. Rep. TS26.267.

[2] Eroupean Commission, "eCall: Time saved = lives saved," Press Release, Brussels, August 21, 2015. website: $http://ec.europa.eu/digital-agenda/en/ecall-time-saved-lives-saved$.

[3] A. Saleem et at. "Four-Dimensional Trellis Coded Modulation for Flexible Optical Communications," *IEEE Journal of Lightwave Technology.,* vol. 35, no. 2, pp. 151-158, Nov. 2017.

[4] C. Studer, C. Benkeser, S. Belfanti, and Q. Huang "Design and Implementation of a Parallel Turbo-Decoder ASIC for 3GPP-LTE," *IEEE Journal of Solid-state Circuits.,* vol. 46, no. 1, pp. 8-17, Jan. 2011.

[5] M. Nader and J. Liu, "Design and implementation of CRC module of eCall in-vehicle system on FPGA," *SAE Technical 2015-01-2844,* 2015, doi:10.4271/2015-01-2844.

[6] M. Nader and J. Liu. "Developing modulator and demodulator for the EU eCall in-vehicle system in FPGAs" in *IEEE, 2016 International Conference on Computing, Networking and Communications (ICNC),* Hawaii, USA, Feb. 15-18, 2016, pp. 1-5.

[7] M. Nader and J. Liu. "FPGA Design and Implementation of Demodulator/Decoder Module for the EU eCall In-Vehicle System" in *International Conference on Embedded Systems and Applications (ESA'15),* Las Vegas, USA, July 27-30, 2015, pp. 3-9.

[8] "Technical Specification Group Radio Access Network; Multiplexing and channel coding (FDD)," 3GPP, Tech. Rep. TS22.212.

[9] D. Viktor, K. Michal, and D. Milan"Impact of trellis termination on performance of turbo codes," in *ELEKTRO,* 2016, pp.48-51.

[10] B. Muralikrishna, G.L. Madhumati, H. Khan, K.G. Deepika, "Reconfigurable system-on-chip design using FPGA," in *2nd International Conference on Devices, Circuits and Systems (ICDCS),* 2014, pp.1-6.

[11] J. J. Rodriguez-Andina, M. J. Moure, and M. D. Valdes, "Features, design tools, and application domains of FPGAs," *IEEE Trans. Ind. Electron.,* vol. 54, no. 4, pp. 18101823, Aug. 2007.

Review on Energy Harvesting For Wireless Sensor Networks

P.Venkatapathi [1] Assistant Professor
Bala tejavath[2] Assistant Professor
Godugu Kiran Kumar[3] Assistant Professor

[1,2]Malla Reddy College of Enginnering
[3]Marri Laxman Reddy Institute of Technology and Management

***ABSTRACT* - Sustainable wireless sensor networks (WSNs) are being widely used nowadays due to two key driving technologies behind them i.e. energy harvesting and energy management.**
Energy harvesting from environmental energy sources such as solar, wind, thermal, mechanical and so forth are introduced from the perspective of energy supply to the WSN, while energy management of WSN such as the design of MAC protocol, design of routing protocol, and dynamic power management technology are presented from the perspective of energy conservation within the WSN itself. To better understand them in details for optimizing the sustainable WSN performance, in this paper, a review of these two enabling technologies are performed. More depth research into their combined efforts for sustainable WSN is presented and then illustrated with a case study. One of the most commonly referred energy harvesting source, i.e. solar energy, and its energy management which includes a new energy forecast model of wireless sensor nodes and a new model of energy distribution in WSNs using data collection protocol is investigated and demonstrated.

I. INTRODUCTION

In recent years, wireless sensor networks are widely used in many areas such as disaster management, infrastructure monitoring, security and surveillance, etc [1]. For these applications, the research works are mostly paying attention to the realization of functions in the designing of wireless sensor networks rather than the sustainability issue of the network. The wireless sensor node always uses power-limited battery as its energy supply. However, there are a number of nodes in the wireless sensor networks and they are always distributed in extensively wide and complex environment, it becomes very difficult to change the battery of wireless sensor nodes on deployment [2]. In order to make wireless sensor networks more practical, researchers began to study the sustainability of the wireless sensor networks, namely, try to extend the life cycle of wireless sensor networks effectively [3]. Energy harvesting and energy management are two key technologies that enable a self-sustainable wireless sensor network.

There are many forms of renewable energy readily available in the environment at which the wireless sensor networks are deployed, such as solar energy, mechanical energy, thermal energy, sound energy, wind power and so on. In this paper, we conduct a review of wide varieties of energy harvesting technologies for wireless sensor networks. We mainly focus on how to transform various forms of energy existing in the environment into electrical energy that can be used to sustain the operations of the wireless sensors. Energy management technology is mainly to solve the problem of energy conservation in wireless sensor networks (WSNs) [4] [5]. Energy management usually includes optimization of medium access and routing protocols, dynamic power management etc. However, if solely relying on reducing energy consumption without energy supplement, it is very difficult to maintain long-term operation of a wireless sensor network.

The objective of this paper is to explore on two key enabling technologies of a self-sustainable and self-autonomous wireless sensor network. Firstly, energy harvesting technologies for wireless sensor network, including solar, wind, sound, vibration, thermal, and electromagnetic are introduced. Secondly, energy management technology used in wireless sensor networks are summarized, which include the design of various MAC protocols, routing protocols, cross layer protocols and dynamic power management technology.

Once the individual energy enhancement technology has been explored and researched, the correlation between both of these key technologies is addressed. To be able to fully optimize the WSN to be self-sustainable, rather than just energy harvesting or improved WSN energy management, it is important to further the research discussion into the combination of both energy harvesting and energy management technologies. A case study on the sustainable wireless sensor network that harvests energy from solar power, the energy model of such wireless sensor networks in green building, and the design of its data collection protocol.

II. ENERGY HARVESTING TECHNOLOGY

As we all know, there are many potential uses of stray energies in our living space, such as solar, wind, heat, mechanical vibration, acoustic, electromagnetic energy. These energy sources are free and pollution free. Much research work on large-scale application of environment energy including solar, wind, geothermal, etc. have already been done and the related technologies are very mature [6]. However, when the problem changes into how to harvest and storage these natural energies in small-scale form to power miniaturized wireless sensor nodes, previous large-scale energy harvesting technologies are no longer applicable. Hence, many research works have been discussed in the literature on this energy harvesting technology for self-sustainable wireless sensor network. Some of the key progresses are described as follows.

A. Solar Energy

For earth, solar energy or light energy is a kind of inexhaustible and clean energy. The basic principle of optical collection is to absorb a large number of photons by the use of photovoltaic materials. If there is enough number of photons to activate the electronic optical pool, electricity can be obtained through appropriate structural design. Because power that can be harvested is greatly depending on the light intensity, optical components are usually placed in an environment with good lighting condition in order to obtain more power. Optical components can be connected in serials to generate the required voltage. As manufacturing cost of optoelectronic components is declining, the selection of solar energy as energy source for wireless sensor networks has become a reasonable technical solution.

The only disadvantage of solar energy is that it is only available during day time (for outdoor environment) or office hour (for indoor environment). A battery is needed to ensure the sensors to be operated all around the clock and the efficiency can be low on cloudy days when sun exposure is very low. A number of recent solar energy harvesting prototypes [7]-[9] for sustainable wireless sensor network are presented in Figure 1.

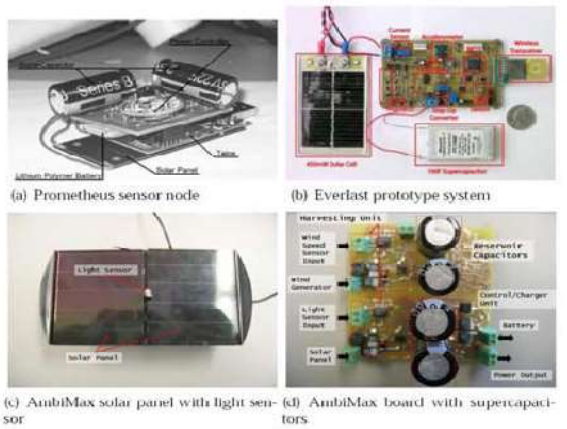


Figure 1: Examples of solar energy harvesting system [7] - [9]

B. Wind Energy

Like any of the commonly available renewable energy sources, wind energy harvesting has been widely researched for high power applications where large wind turbinegenerators (WTGs) are used for supplying power to remote loads and grid-connected applications. Although very few research works are reported in the literature on small-scale wind energy harvesting, some efforts to generate power at a very small-scale have been made recently [10]-[12], and some are presented in Figure 2.

The main disadvantage regarding wind power is unreliability factor, as the strength of the wind is not constant and unpredictable, hence it does

not produce the same amount of electricity all the time. In addition, since it involves moving mechanical part, it can be noisy.

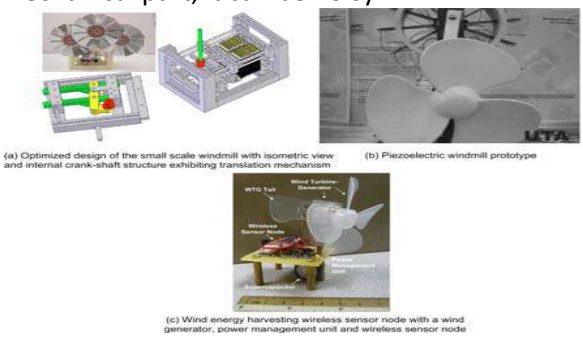


Figure 2: Examples of wind energy harvesting system [10]-[12]

C. Thermal Energy

Research on thermoelectric technology began in 1940's, reached its peak in 1960's. And this technology was successfully used on the spacecraft. Temperature difference generator is featured with characteristics such as small, light weight, no vibration, no noise, less maintenance and can work for long hours under harsh environment. It is suitable to act as low power less than 5W and usually mounted in a variety of unmanned surveillance sensors, tiny short-range communication devices, and medical instrumentation. At present, the relevant products have been widely used. German scientists have invented a new type of battery using the temperature of human body to produce electricity, which can provide long-term "power" for portable miniature electronic devices and eliminates the trouble of charging or replacing batteries. For example, temperature difference which equal to 5 degree C between human skin and clothes can be took advantage of and provide sufficient energy for a common watch. Some examples of the thermal energy harvesting systems are presented in Figure 3.

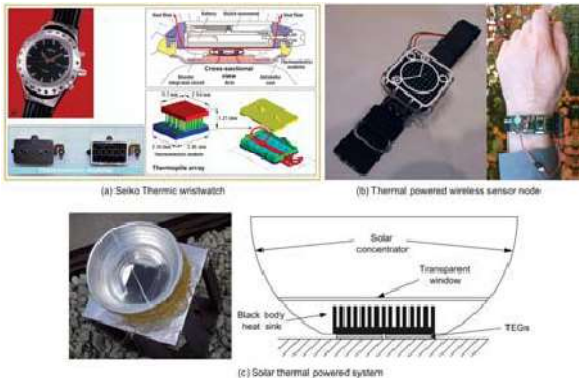


Figure 3: Examples of thermal energy harvesting system [13] -[15]

V. CONCLUSION

In this paper, we provide a comprehensive review on some common energy harvesting technologies of wireless sensor networks, and the introduction of energy management technology. We demonstrate an example of sustainable wireless sensor networks based on solar energy which is for green building. The challenge to harvest environment energy is discussed.

REFERENCES

[1] Y.K. Tan and S.K. Panda (2010). Review of Energy Harvesting Technologies for Sustainable WSN, Sustainable Wireless Sensor Network, Winston Seah (Editor) and Yen Kheng Tan (Editor-in-Chief), ISBN 978-953-307-297-5, INTECH. Available from: http://www.intechopen.com/articles/show/title/review-of-energyharvesting- technologies-for-sustainable-wsn

[2] C. Alippi, G. Anastasi, M. Di Francesco, M. Roveri, “Energy management in wireless sensor networks with energy-hungry sensors”, *IEEE Instrumentation & Measurement Magazine*, vol.12, issue.2, pp.16- 23, 2009.

[3] V. Raghunathan, S. Ganeriwal, M. Srivastava, “Emerging techniques for long lived wireless sensor networks”, *IEEE Communications Magazine*, vol.44, issue.4, pp.108-114, 2006.

[4] C.Y. Chong, S.P. Kumar, “Sensor networks: Evolution, opportunities, and challenges”, *Proceedings of the IEEE Sensor Networks and Applications*, vol.91, no.8, pp.1247-1256, 2003.

[5] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci, "A Survey on Sensor Networks", *IEEE Communications Magazine*, 2002.

[6] Y.K. Tan and S.K. Panda, "Optimized Wind Energy Harvesting System Using Resistance Emulator and Active Rectifier for Wireless Sensor Nodes", *IEEE Transactions on Power Electronics*, vol.26, issue.1, pp.38- 50, 2011.

[7] X.F. Jiang, J. Polastre and D.E. Culler, "Perpetual environmentally powered sensor networks", *Fourth International Symposium on Information Processing in Sensor Networks (IPSN)*, pp.463-468, 2005.

[8] F.I. Simjee, P.H. Chou, "Efficient charging of supercapacitors for extended lifetime of wireless sensor nodes", *IEEE Transactions on Power Electronics*, vol.23, issue.3, pp.1526-1536, 2008.

[9] C. Park, P.H. Chou, "AmbiMax: Autonomous Energy Harvesting Platform for Multi-Supply Wireless Sensor Nodes", *3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks (SECON)*, vol.1, pp.168-177, 2006.

[10] S.Priya, C.T. Chen, Darren Fye, Jeff Zahnd, "Piezoelectric Windmill: A Novel Solution to Remote Sensing", *Japanese Journal of Applied Physics*, vol.44, no.3, pp.L104L107, 2005.

[11] R. Myers, M. Vickers, Kim Hyeoungwoo, S. Priya, "Small scale windmil", *Applied Physics Letters*, vol.90, no.5, p.54106-1-3, 2007.

[12] Y.K. Tan and S.K. Panda, "Self-Autonomous Wireless Sensor Nodes with Wind Energy Harvesting for Remote Sensing of Wind-Driven Wildfire Spread", *IEEE Transactions on Instrumentation and Measurement*, 2011.

[13] Nicholas S. Hudak and Glenn G. Amatucci, "Small-scale energy harvesting through thermoelectric, vibration, and radio frequency power conversion", *Journal of Applied Physics*, vol.103, no.10, pp.101301(1- 24), 2008.

[14] V. Leonov, T. Torfs, P. Fiorini, C. Van Hoof, "Thermoelectric Converters of Human Warmth for Self-Powered Wireless Sensor Nodes", *IEEE Sensors Journal*, vol.7, issue.5, pp.650-657, 2007.

[15] H.A. Sodano, G.E. Simmers, R. Dereux, D.J. Inman, "Recharging batteries using energy harvested from thermal gradients", Journal of Intelligent Material Systems and Structures, vol.18, no.1, pp.3-10, 2007.

A Simple Yet Efficient Accuracy Configurable Adder Design

Vavaldas Harshini,M.tech scholar,Dept of ECE,MRCE, Hyderabad,harshiniraman143@gmail.com

***Abstract*—Approximate computing is a promising approach for low power IC design and has recently received considerable research attention. To accommodate dynamic levels of approximation, a few accuracy configurable adder designs have been developed in the past. However, these designs tend to incur large area overheads as they rely on either redundant computing or complicated carry prediction. Some of these designs include error detection and correction circuitry, which further increases area. In this work, we investigate a simple accuracy configurable adder design that contains no redundancy or error detection/correction circuitry and uses very simple carry prediction. Simulation results show that our design dominates the latest previous work on accuracy-delay-power tradeoff while using 39% lower area. In the best case the iso-delay power of our design is only 16% of accurate adder regardless of degradation in accuracy. One variant of this design provides finer-grained and largertunability than the previous works. Moreover, we propose a delay-adaptive self-configuration technique to further improve accuracy-delay-power tradeoff. The advantages of our method are confirmed by theapplicationsinmultiplicationandDCTcomputing.**

***Index Terms*—Approximate computing, accuracy configurable adder, delay-adaptive reconfiguration, low power design.**

I. Introduction

POWER constraints are a well-known challenge in advanced VLSI technologies. Low power techniques for the conventional exact computing paradigm have been already extensively studied. A comparatively new direction is approximate computing, where errors are intentionally allowed in exchange for power reduction. In many applications, such as audio, video, haptic processing and machine learning, occasional small errors are indeed acceptable. Such error-tolerant applications are found in abundance in emerging applications and technologies.

A great deal of approximate computing research has been concentrated on arithmetic circuits, which are essential building blocks for most of computing hardware. In particular, several approximate adder designs have been developed [1]–[14]. One such design [2] achieves 60% power reduction for DCT (Discrete Cosine Transform) computation without making any discernible difference to the images being processed. In realistic practice, accuracy requirements may vary for different applications. In mobile computing devices, different powermodesmayentaildifferentaccuracyconstraintsevenfor the same application. Specifically, arithmetic accuracy can be adjusted at runtime using methods such as dynamic voltage and frequency scaling (DVFS) to obtain the best accuracy-power tradeoff. The benefit of runtime accuracy adjustment is demonstrated in [3], but their approximation is realized by voltage over-scaling, where errors mostly occur at the timing-critical path associated with the most significant bits, i.e., errors are oftenlarge.

To reduce the overall error, a few approximate designs have been developed by intentionally allowing errors in lower bits with shorter carry chain in addition operation. In [4], a design that considers only the previous *k* inputs instead of all input bits can approximate the result with the benefit in half of the logarithmic delay. Reliable variable latency carry select adder (VLCSA) shows a speculation technique which introduces carry chain truncation and carry select addition as a basis [7]. A series of Error Tolerant Adders (ETAI, ETAII, ETAIIM), which truncate the carry propagation chain by dividing the adder into several segments, have been proposed [8]–[10]. Correlation-aware speculative adder (CASA) in [11] relies on the correlation between MSBs of input data and carry-in values. Another approximate adder that exploits the generate signals for carry speculation is presented [12]. These designs focus on static approximation which pursues almost correct results at the required accuracy. However, in some applications such as image processing or audio/video compression, the required accuracy might vary during runtime. To meet the need of runtime accuracy adjustment, a series of designs are developed to implement accuracy configurable approximation whichcouldbereconfiguredonlinetosavemorepower.

A few accuracy configurable adder designs that use approximation schemes other than voltage over-scaling have been proposed. An early work [15], called ACA, starts with an approximate adder and augments it with an error detection and correction circuit, which can be configured to deliver varying approximation levels or accurate computing. Its baseline approximate adder contains significant redundancy and the error detection/correction circuit further increases area overhead. The ACA design [15] is generalized to a flexible framework GeAr in [16]. In both ACA and GeAr, the error correction must start from the least significant bits and hence accuracy improves slowly in the progression ofconfigurations. The work of Accurus [17] modifies ACA/GeAr to overcome this drawback and achieves graceful degradation. However, in ACA, GeAr as well as Accurus, the error correction circuit is pipelined, implying that the computation in accurate mode takesmultipleclockcyclesandcausesdatastalls.

This work was partially supported by NSF (CCF-1255193, CCF-1525749 and CCF-1525925).

W. Xu and J. Hu are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, 77843 USA (e-mail: wbxu@tamu.edu; jianghu@tamu.edu).

S. S. Sapatnekar is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, 55455 USA (email: sachin@umn.edu).

An alternative direction of accuracy configurable adder design is represented by GDA [18] and RAP-CLA [19]. These methods start with an accurate adder and use carry prediction for optional approximation. As such, they no longer need error detection/correction and do not incur any data stall. In addition, they intrinsically support graceful degradation. The GDA design [18] is composed by accurate CRA (Carry Ripple Adder) and extra configurable carry prediction circuitry, similar as the carry look-ahead part of CLA (Carry Look-ahead Adder). Thus, its area is generally quite large. RAP-CLA [19] is based on accurate CLA design and reuses a portion of the carry look-ahead circuit as carry prediction. This leads to an overall area that is less than GDA but greater than CLA. In [19], the carry-prediction-based approach is shown to be superior to error-correction-based design[16].

In this paper, we propose a new carry-prediction-based accuracy configurable adder design: **SARA** (Simple Accuracy Reconfigurable Adder). It is a simple design with significantly less area than CLA, which, to the best of our knowledge, has not been achieved in the past in accuracy configurable adders. SARA inherits the advantages of all previouscarry-prediction-basedapproaches:noerrorcorrectionoverhead,nodatastall andallowinggracefuldegradation.ComparedtoGDA[18], SARA incurs 50% less PDP (Power Delay Product) and can reach the same PSNR (Peak Signal-to-Noise Ratio). Moreover, SARA demonstrates remarkably better accuracy-power-delay tradeoff than the latest, and arguably the best, previous work RAP-CLA [19]. A delay-adaptive reconfiguration technique is developed to further improve the accuracy-power-delay tradeoff. The proposed designs are also validated by multiplication and DCT computation in image processing.

II. Prior Works and Rationale of Our Design

Wereviewafewrepresentativeworksonaccuracyconfigurable adder design and show the relation with our method. These designs can be generally categorized into two groups: error-correction-based configurations [15]–[17] and carry-prediction-based configurations [18], [19].

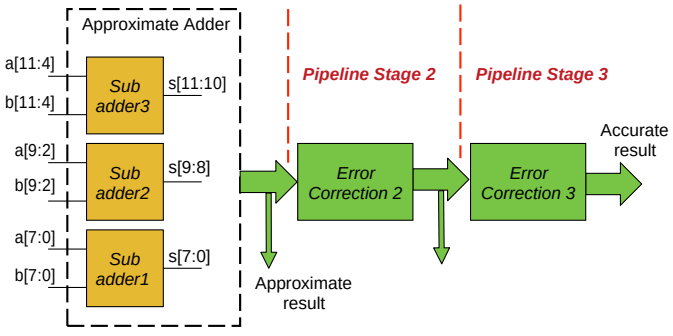


Fig. 1. Error-correction-based configurable adder.

Themainideaofanerror-correction-basedapproach[15]–[17] is shown in Figure 1. The scheme starts with an approximate adder (the dashed box), where the carry chain is shortened by using separated sub-adders with truncated carry-in.Inordertoreducethetruncationerror,thebit-width in some sub-adders contains redundancy. For example, *sub-adder2* calculates the sum for only bit 8 and 9, but it is an 8-bit adder using bit $[9:2]$ of the addends, 6 bits of which are redundant. Even with the redundancy, there is still residual error which is detected and corrected by additional circuits. In Figure 1, the errors of *sub-adder2* must be corrected by *error-correction2* before the errors of *sub-adder3* are rectified by *error-correction3*. As such, the configuration progression always starts with small accuracy improvements. The redundancy and error detection/correction incur large area overhead. Since the error correction circuits are usually pipelined, an accuratecomputationmaytakemultipleclockcyclesandcould stallentiredatapath,dependingontheaddendvalues.

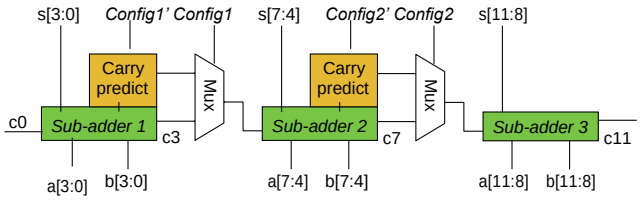


Fig. 2. Carry-prediction-based configurable adder.

Theframeworkofcarry-prediction-basedmethods[18],[19] is shown in Figure 2. These schemes start with an accurate adder design, which is formed by chaining a set of sub- adders. Each sub-adder comes with a fast but approximated carry prediction circuit. By selecting between the carry-out from sub-adder or carry prediction, the overall accuracy can be configured to different levels. Such an approach does not need error detection/correction circuitry. Moreover, the configuration of higher bits is independent of lower bits. This leads to fast convergence or graceful degradation in the progression of configurations. In GDA [18], the sub-adders are CRA designs while the carry-prediction circuit is similar tothe carry look-ahead part of CLA. Further, its carry prediction can be configured to different accuracy levels. However, the complicated carry prediction induces large area overhead. The RAP-CLA scheme [19] uses CLA for its baseline where the carry-ahead of each bit is computed directly from theaddends of all of its lower bits. Its carry prediction reuses a part of the look-ahead circuit rather than building extra dedicated prediction circuitry, and hence is more area-efficient than GDA.ButitsbaselineismuchmoreexpensivethanGDA.

TABLE I
Comparison of characteristics for different techniques.

Method	Baseline sub-adder	Error correction	Graceful degradation	Carry prediction
ACA [15]	Redundant CRA	Yes	No	No
GeAr [16]	Redundant CRA	Yes	No	No
Accurus [17]	Redundant CRA	Yes	Yes	No
GDA [18]	CRA	No	Yes	Stand-alone
RAP-CLA [19]	CLA	No	Yes	Reuse
SARA (ours)	CRA	No	Yes	Reuse

Our design is a carry-prediction-based approach. Its sub-adders are CRA instead of expensive CLA as in RAP-CLA. Itscarrypredictionalsoreusespartofthesub-addersrather

than having dedicated prediction circuitry. As such, it avoids the disadvantages of both GDA and RAP-CLA. A comparison among the characteristics of these different techniques is provided in Table I.

III. Simple Accuracy Reconfigurable Adder

A. Preliminaries

An N-bit adder operates on two addends $A = (a_N, a_{N-1}, ..., a_i, ..., a_1)$ and $B = (b_N, b_{N-1}, ..., b_i, ..., b_1)$. For bit i, its carry-in is c_{i-1} and its carry-out is c_i. Defining the carry generate bit $g_i = a_i b_i$, propagate bit $p_i = a_i \oplus b_i$ and kill bit $k_i = \bar{a}_i \bar{b}_i$, the conventional full adder computes the sum s_i and carry c_i according to

$$s_i = p_i \oplus c_{i-1}, \tag{1}$$

$$c_i = g_i + p_i \cdot c_{i-1}. \tag{2}$$

A gate level schematic of conventional full adder is provided in Figure 3(a). A CRA is used to chain N bits of conventional full adders together.

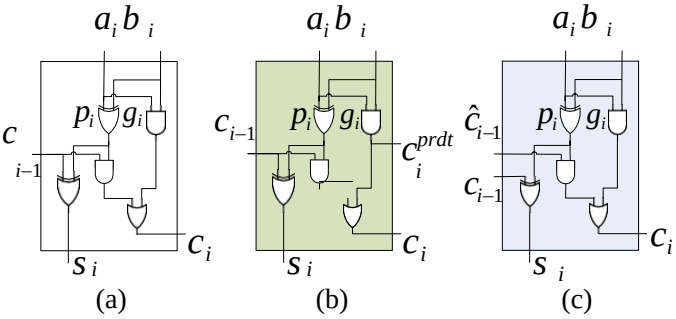


Fig.3. (a) Conventional full adder; (b) Our carry-out selectable full adder; (c) Our carry-in configurable full adder.

By applying Equation (2) recursively, one can get

$$c_i = g_i + p_i g_{i-1} + ... + g_1 \prod_{k=2}^{i} p_k + c_0 \prod_{k=1}^{i} p_k. \tag{3}$$

This equation implies that c_i can be computed directly from g and p of all bits, without waiting for the c of its lower bits to be computed. This observation is the basis for CLA adder.

B. SARA: Simple Accuracy Reconfigurable Adder Design

In SARA, an N-bit adder is composed by K segments of L-bit sub-adders, where $K = \lfloor N/L \rfloor$ (see Figure 2). Each sub-adder is almost the same as CRA except that the MSB (Most Significant Bit) of a sub-adder, which is bit i, provides a carry prediction as

$$c_i^{prdt} = g_i \tag{4}$$

For the LSB (Least Significant Bit) of the higher-bit sub-adder, which is bit $i+1$, its carry-out c_{i+1} can be computed using one of two options: either by the conventional $c_{i+1} = g_{i+1} + p_{i+1} \cdot c_i$, or by using the carry prediction as

$$c_{i+1} = g_{i+1} + p_{i+1} \cdot c_i^{prdt} = g_{i+1} + p_{i+1} \cdot g_i \tag{5}$$

The selection between the two options is realized using MUXes as in Figure 4 and the MUX selection result is denoted as $\hat{c}_i$. Comparing Equation (5) with (3), we can see that the carry prediction is a truncation-based approximation to carry computation[1]. Therefore, $\hat{c}_i$ can be configured to either accurate mode or approximation mode, i.e.,

$$\hat{c}_i \leftarrow \begin{cases} c_i^{prdt}, & \text{if approximation mode} \\ c_i, & \text{if accurate mode.} \end{cases} \tag{6}$$

It should be noted that the carry prediction c_i^{prdt} reuses g_i in an existing full adder instead of introducing an additional dedicated circuit as in [18] or Figure 2. This prediction scheme makes a very simple modification to the conventional full adder, as shown in Figure 3(b).

One can connect $\hat{c}_i$ to its higher bit $i+1$ to compute both carry c_{i+1} and sum s_{i+1}, as in GDA [18] and RAP-CLA [19]. We suggest an improvement over this approach by another simple change as in Figure 3(c), where s_{i+1} is based on c_i instead of $\hat{c}_i$. Such approach can help reduce the error rate in outputs when an incorrect carry is propagated. Because the sum keeps accurate and the carry will not be propagated when addends are exactly the same. Moreover, out of all four configurations of sum/carry calculation by approximate/accurate carry-in, the most meaningful way is to have sum bit calculated by accurate carry and make carry bit configurable. So sum s_{i+1} is calculated directly by accurate carry c_i without the option of c_i^{prdt}. Applying this in SARA as in Figure 4, in the approximation mode, computing s_{j+1} from c_j can still limit the critical path to be between c_i^{prdt} and s_{j+1}, but has higher accuracy than computing s_{j+1} from $\hat{c}_j$. Compared to sum computation in GDA and RAP-CLA, this technique improves accuracy with almost no additional overhead. Compared to CRA, the overhead of SARA is merely the MUXes, which is almost the minimum possible for configurable adders.

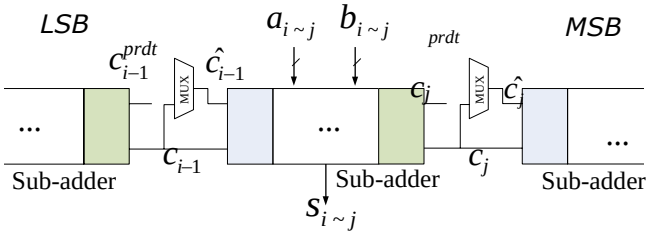


Fig. 4. Design of SARA.

Although s_{j+1} is calculated by accurate carry c_j, its delay can still be reduced by approximate carry in lower sub-adder. In a multi-bit adder, the delay of sum bit depends on the carry chain propagated from its lower bits. In our SARA structure, even when accurate carry c_j is propagated at bit j, the carry chain might be truncated by approximate carry in other lower bits. In Figure 4, when c_{i-1}^{prdt} is propagated, the delay of s_{j+1} is reduced as its path is shorten to be between bit $i-1$ and

[1] A similar approximation is used in static approximate adder design [12].

$j+1$. We can take the 12-bit adder in Figure 5 as an example. For 12-bit SARA working in approximate mode, the sum $s_9$uses the accurate carry $c_8$from a lower sub-adder (bit 5 to 8). But $c_8$is propagated from approximate carry $c_4^{prdt}$of another sub-adder (bit 1 to 4). As shown in the figure, the delay of $s_9$in SARA is about 6 stages. Compared with the same bit in CRA, the delay of sum bit $s_9$in SARA is reduced by 3 stages. Similar delay reduction can be observed in other sum bits (bit 6 to 12). For sums at bit 1 to 5, their delay is the same as CRA because they are using an accurate carry $c_0$from LSB. As a result, the maximum delay in 12-bit SARA is reduced, since for a multi-bit adder its maximum delay depends on the longest criticalpath.

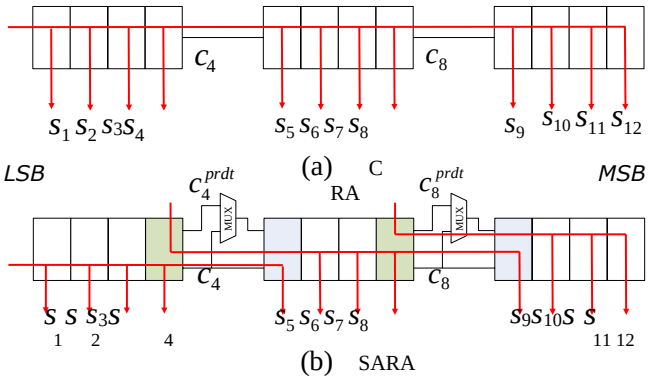


Fig. 5. Implementation of 12-bit adder in (a) CRA and (b) SARA.

C. Usage ofSARA

When$\hat{c}_i$isconfiguredtobe$c_i$forall$K$sub-adders,SARA operates very much like the CRA, where the critical path is along$N$-bitfulladders.If all$\hat{c}_i$areselectedtobec_i^{prdt},the critical path is shortened to roughly L-bit full adders. This large delay reduction can be translated to power reduction by supply voltage scaling. Voltage scaling (reducing supply voltage) on digital circuits will lead to increase in delay. So wecanreducethesupplyvoltageonSARAtomakeitscritical delay same as that of CRA under normal voltage. As the supply voltage decreases, the power consumption could be reduced.Therecan be2^{K-1}differentconfigurations.Fortwo configurations with the same critical path length, obviously we only need the one with higher accuracy. Therefore,there are$K$effectiveconfigurations,withcriticalpathlengthof$L$- bit, $2L$-bit, ..., $KL\cdot N$-bit full adders. The delay of such configurabledesignvariesaccordingtoconfiguredaccuracy, whichresultsindifferentpowerreductionbyvoltagescaling.

IV. SARA ErrorAnalysis

Inthissection,wegiveatheoreticalanalysisontheexpected errorofourSARAdesignandvalidatetheresultsbynumerical experiments. To make it easier for readers to follow the analysis, we list the parameters used in this section as TableII.

TABLE II
DEFINITION OF PARAMETERS FOR ERROR ANALYSIS

Parameter	Definition
p_i	propagate bit at bit i
g_i	generate bit at bit i
k_i	kill bit at bit i
c_i	accurate carry-out bit at bit i
c_i^{prdt}	approximate carry-out bit at bit i
$\hat{c}_i$	carry-in bit at bit $i+1$
ER_i^{prdt}	error rate of c_i^{prdt}
$\widehat{ER}_i$	errorrateof$\hat{c}_i$

For any bit i in carry-out selectable full adder as in Figure3(b),anerrorinapproximatecarry-outoccurswhen $c_i^{prdt} \neq c_i$. There is only one situation where this error may happen: when $c_{i-1}=1$, $p_i=1$, $c_i^{prdt}=0$ and $c_i=1$. Then the error rate, or probability of such error, is given by

$$\begin{aligned} ER_i^{prdt} &= P(c_i^{prdt} \neq c_i) = P(c_i^{prdt}=0, c_i=1) \\ &= P(c_{i-1}=1, p_i=1) \\ &= P(c_{i-1}=1)P(p_i=1) \end{aligned} \tag{7}$$

where P indicates probability and the last part assumes that $c_{i-1}$and $p_i$are independent of each other. Then, if the approximate/accurate carry-out can be selected by a MUX gate,theerrorrateofMUXoutput$\hat{c}_i$is

$$\widehat{ER}_i = P(\hat{c}_i \neq c_i) = \begin{cases} ER_i^{prdt}, & \text{if } \hat{c}_i \leftarrow c_i^{prdt} \\ 0, & \text{if } \hat{c}_i \leftarrow c_i. \end{cases} \tag{8}$$

Let'sconsideraconfigurationofSARAinFigure4,which has both bit j and bit $i+1$ in approximate mode. For the sub-adder which calculates addends from bit i to bit j, its LSB(bit$i$)isusingcarry-inconfigurablefulladder,whileits MSB(bit$j$)isincarry-outselectablefulladder.Accordingto Equation(7)and(8),theerrorrateof$\hat{c}_j$isdeterminedbythe probabilities of $c_{j-1}=1$ and $p_j=1$.

$$\widehat{ER}_j = P(c_{j-1}=1)P(p_j=1) \tag{9}$$

Accordingtothelogicofaddition,thecarry-outbitiscalculatedbythecarry-inandaddends.Therearetwocaseswhich can result in $c_{j-1}=1$: generate bit g_{j-1}should be 1 in case of carry-in $c_{j-2}=0$; or kill bit k_{j-1}must be 0 when carry-in comes with $c_{j-2}=1$. Then, the probability of $c_{j-1}=1$ can be computed by the probability of $c_{j-2}=1$ as

$$\begin{aligned} P(c_{j-1}=1) &= P(c_{j-2}=0, g_{j-1}=1) + P(c_{j-2}=1, k_{j-1}=0) \\ &= P(c_{j-2}=0)P(g_{j-1}=1) + P(c_{j-2}=1)P(k_{j-1}=0) \\ &= [1-P(c_{j-2}=1)]P(g_{j-1}=1) + P(c_{j-2}=1)P(k_{j-1}=0). \end{aligned} \tag{10}$$

Similarly,theprobabilityof$c_{j-2}=1, c_{j-3}=1, \ldots, c_{i+1}=1$ canbecalculatedusingthesameformula.Fortheprobability of$c_i=1$,it'salittledifferentbecausethecarry-out$c_i$inour carry-inconfigurablefulladderisbasedonpredictedcarry-in $\hat{c}_{i-1}$insteadofc_{i-1}.Consideringthatbit$i$1isconfigured in approximate mode, wehave

$$P(\hat{c}_{i-1}=1) = P(c_{i-1}^{prdt}=1) = P(g_{i-1}=1). \tag{11}$$

Then, the probability of $c_i=1$ can be expressed as

$$P(c_i=1) = [1 - P(g_{i-1}=1)]P(g_i=1) + P(g_{i-1}=1)P(k_i=0). \tag{12}$$

TABLE III
ERROR RATE OF SUB-ADDER WITH DIFFERENT WIDTH

Sub-adder length L	Calculated error rate	Simulated error rate
1	1/8 = 0.125	0.1257
2	3/16 = 0.1875	0.1879
3	7/32 = 0.21875	0.2187
4	15/64 = 0.234375	0.2347
5	31/128 = 0.2421875	0.2424
6	63/256 = 0.24609375	0.2464

By expanding Equation (10) recursively till bit i, the probability of $c_{j-1}=1$ can be calculated by a function of generate bit and kill bit from bit $i-1$ to bit $j-1$.

$$P(c_{j-1}=1)=f\{P(g_{i-1}=1),...,P(g_{j-1}=1), P(k_i=0),...,P(k_{j-1}=0)\}. \quad (13)$$

Assuming that the inputs for adder are uniformly distributed random numbers, we have $P(g=1)=1/4, P(k=0)=3/4$. As the length of sub-adders varies from 1 to 6, the error rates of $\hat{c}_j$ calculated by Equation (9) are listed in the second column of Table III. Corresponding data from numerical simulation in Matlab are also presented in the last column. The error rates calculated by our method match well with experiment results, which demonstrates the correctness of our mathematical analysis. We can also observe that as the length of sub-adder increases the error rate is bounded by 0.25. That is because when the length of sub-adder comes to infinite the probability of $c=1$ will become 0.5 as the normal carry in accurate adder.

Theorem 1. *If* $\mathbf{I}$ *is the set of bits with MUX at output, the expected error of SARA for unsigned integers is*

$$\sum_{\forall i\in \mathbf{I}} \widehat{ER}_i \cdot P(p_{i+1}=1)\cdot 2^{i+1}.$$

Proof. The overall expected error of SARA can be calculated by summing respective error introduced by every approximate bit from LSBs to MSBs. But the propagation of inaccurate carry bit may cause error in higher bit which also be counted in the calculation of lower bit. So we need to exclude those errors to avoid over-calculation in the total error.

Let's consider the SARA design in Figure 4 which have approximate configuration at both bit $i-1$ and bit j. Assuming that bit $i-1$ is the lowest bit configured in approximate mode, we know that all sum bits $s_k (k \in [1, i-1])$ as well as carry bit c_{i-1} are accurate.

$$c_{i-1} = c_{i-1}^{acc} \quad (14)$$

Then the probability that carry prediction at MUX output $\hat{c}_{i-1}$ mismatches with accurate carry c^{acc}_{i-1} should be the same as the error rate of max ouput $\hat{c}_{i-1}$.

$$P(\hat{c}_{i-1} \neq c_{i-1}^{acc}) = P(\hat{c}_{i-1} \neq c_{i-1}) = \widehat{ER}_{i-1} \quad (15)$$

According to the structure of carry-in configurable full adder (Figure 3(c)), sum bit s_i calculated from c_{i-1} is always accurate; however, the carry-out bit c_i becomes conditionally accurate which depends on both carry-in bit and propagate bit. As shown in Equation (16), the scenario of accurate carry-out can be attributed to two conditions: when the carry-in is not accurate, the carry-out bit becomes accurate as the propagate bit is false; otherwise, it must be accurate no matter what kind of addends are given.

$$P(c_i=c_i^{acc}) = P(\hat{c}_{i-1}=c_{i-1}^{acc}) + P(\hat{c}_{i-1} \neq c_{i-1}^{acc}) P(p_i=0) \quad (16)$$

Its complementary part, the probability of inaccurate carry c_i, can be expressed as

$$\begin{aligned} P(c_i \neq c_i^{acc}) &= P(\hat{c}_{i-1} \neq c_{i-1}^{acc}) P(p_i=1) \\ &= P(\hat{c}_{i-1} \neq c_{i-1}) P(p_i=1) \\ &= \widehat{ER}_{i-1} \cdot P(p_i=1). \end{aligned} \quad (17)$$

As a result, the approximation at bit $i-1$ would cause an inaccurate carry-in c_i at bit $i+1$, which introduces the magnitude of 2^i to the overall error in final result. Then the expected error introduced by approximation at bit $i-1$ can be estimated by

$$E[e_{i-1}] = P(c_i \neq c_i^{acc}) \cdot 2^i = \widehat{ER}_{i-1} \cdot P(p_i=1) \cdot 2^i. \quad (18)$$

Next, we consider the expected error introduced by approximation at bit j. As bit j is not the lowest bit in approximate mode, there is a chance that the propagation of inaccurate carry from bit $i-1$ induces error at bit j while it has be taken into account in the error calculation of bit $i+1$. Then the problem is whether the carry c_j is accurate when there is a mismatch between $\hat{c}_j$ and c_j. If not, we need to exclude the impact from lower bit when estimating the error at bit j. Let's answer this question in the following cases.

- Case 1: If any propagate bit in sub-adder (bit i to j) equals 0, the error propagation by inaccurate carry will be paused. In another word, the error carried by inaccurate carry bit cannot be propagated to higher bit any more, because the carry-out is independent of carry-in when propagate bit is false. In this case, the carry c_j should be always accurate regardless of the configuration at bit j.
- Case 2: If all propagate bits of sub-adder equal 1, the value of inaccurate carry $\hat{c}_{i-1}$ (0 instead of 1) will be propagated to c_j. In this situation, the actual value of c_j propagated from bit $i+1$ must be 0, while the accurate value should be 1. Assuming that $\hat{c}_j$ mismatches with c_j, we can state that the value of $\hat{c}_j$ must be 1. However, it conflicts with the generation of c_j, because carry c_j is the logical conjunction of $\hat{c}_j$ and $p_j c_{j-1}$. So there should be no mismatch between $\hat{c}_j$ and c_j in this case.

In conclusion, when there is a mismatch between $\hat{c}_j$ and c_j, the value of carry c_j must be accurate. We can further conclude that the contributions of every approximate bit to the total error are independent to each other. Similar to bit $i-1$, the expected error at bit j can be estimated by

$$E[e_j] = \widehat{ER}_j \cdot P(p_{j+1}=1) \cdot 2^{j+1}. \quad (19)$$

Thus, the total error can be obtained by summing up the errors respectively introduced by every approximate bit.

$$E = \sum_{\forall i\in \mathbf{I}} E[e_i] = \sum_{\forall i\in \mathbf{I}} \widehat{ER}_i \cdot P(p_{i+1}=1) \cdot 2^{i+1} \quad (20)$$

□

If input addends are random variables following uniform distribution, the expected error of SARA is given by

$$E = \sum_{\forall i \in I} \widehat{ER}_i \cdot 2^i \qquad (21)$$

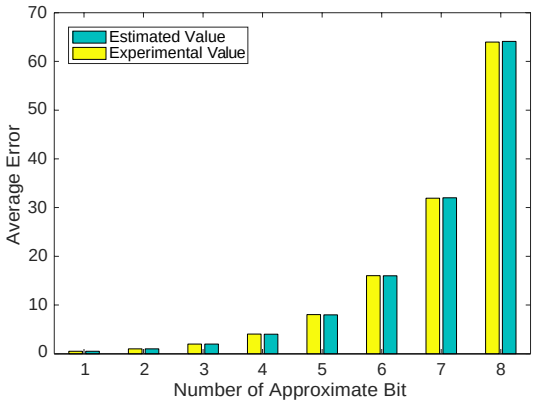


Fig. 6. Average error of 9-bit SARA in different configuration.

We can verify Equation (21) by numerical simulation of a 9-bit SARA design. In our experiment, SARA consists of 9 sub-adders whose width is 1 bit. The results are from 200Krun of Monte Carlo simulation with uniform distributed numbers as input. As shown in Figure 6, there are 2 sets of data for comparison, experimental data are obtained directly in experiments and estimated data are calculated by Equation (21). The average errors from experiment are almost the same asthe estimated values. According to the analysis above, we can estimatetheaverageerrorofSARAinanyconfiguration,given the distribution of inputnumbers.

Since $\exists |K$ 1, the error of the worst case approximation modeincreaseswiththenumberofsub-adders,K.Inaddition, areaoverheadincreaseswith$K$.Ontheotherhand,alarge Kimpliessmaller$L$,andthusoftenfacilitatesshortercritical path and more power reductions. Therefore, K significantly affects the tradeoff among accuracy, power, delay and area.

V. Delay-Adaptive Reconfiguration of SARA

Almost all previous works on accuracyconfigurable adder [15]–[19] reasonably assume that accuracy configuration is decided by architecture/system level applications. Wepropose a self-configuration technique for the scenarios wherearchitecture/system level choice is either unclear or difficult.Simulation results show that SARA with the self-configurationoutperforms several previous static approximate adder designs.The main idea of self-configuration is based on the observa-tion that the actual worst case path delay depends on addendvalues. Specifically, the actual path delay is large only whena carry is propagated through several consecutive bits. Any false propagate bit from the addends results in a shorter carry propagation chain. When the actual carry propagation chain is short, there is no need to use approximation configuration, which is intended to cut carry chain shorter. We propose a Delay Adaptive Reconfiguration (**DAR**) technique: the output of a MUX in SARA is set to approximation mode only when a potentially long carry chain is detected. Compared to the constantly-approximate configuration, some errors for actual short carry chains are avoided, the actual long carry chain is cut shorter, and delay/power reduction can be still obtained.

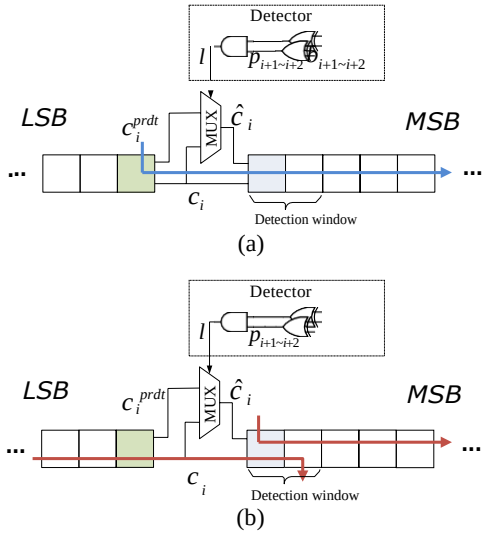


Fig. 7. Design and operation of delay-adaptive reconfiguration for SARA.

The long carry chain detection and SARA-DAR design are shown in Figure 7(a). When MUX is switched to accurate mode by any false propagate bit in detection window, the actual carry chain is retained by the position of false propagate bit. To obtain a shorter carry chain in accurate mode, the detection window for MUX at bit i in MSB should start from bit i+1. In the example of Figure 7, we use a detectionwindow of 2 bits ($p_{i+1}$and p_{i+2}) to tell if there is a carry propagation acrosstwosub-adders,andconfiguretheMUXaccordingto

$$\hat{c}_i \leftarrow \begin{cases} c_i^{prdt}, & \text{if } p_{i+1} \cdot p_{i+2} \text{ is true} \\ c_i, & \text{otherwise.} \end{cases} \qquad (22)$$

Inapproximationmode,theeffectivecarrychainisrepresentedbythe bluelineinFigure7(a)anditslengthisnogreaterthan $L + 1$ bits. When the MUX is set to accuratemode,thecarrychainisindicatedbytheredlinesinFigure 7(b)andtheirlengthscanberestrainedtowithin$L$+2bits.Sincethep ropagate bits only depend on local primary inputs,wecanreuse propagate bits in higher bits to save cost. Notethatinthiscasethedetectionoverheadhereisalmostthemini mumpossible,i.e.,onlyoneNANDgateforconfiguringeachMU X.In Figure 7, we use 2-bit detection window, whichcanbegeneralizedto$W$-bit.Then,theerrorrateforMUXatbit$i$ becomes

$$\widehat{ER}_i^{dar} = ER_i^{prdt} \cdot \prod_{j=1}^{W} P(p_{i+j} = 1) \qquad (23)$$

Thedetectionwindowsize$W$decidesthetradeoffbetween accuracyandtheeffectivecarrychainlengthinaccuratemode, which is $L + W$. When W increases, the error ratedecreases while the critical path length in accurate modeincreases.

VI. Experimental Results

A. Experiment Setup and Evaluation

Our SARA, SARA-DAR and several previous designs are synthesized to 32-bit adders by Synopsys Design Compiler using the Nangate 45nm Open Cell Library. The synthesized circuits are placed and routed by Cadence Encounter. The default supply voltage level is 1.25V. To make fair comparisons across architectures, we describe all designs bystructural modeling in Verilog to reduce the impact of synthesis and optimization. For comparison, we synthesize the accurate adder in behavioral modeling which is described by expressional operator in Verilog. The netlist of such accurate adder should be automatically optimized by synthesizer in Design Complier, which is different from any man-craft gate-level design. In addition, we set the same supply voltage and no delayconstraintonalldesignsforthesamereason.

The evaluation of accuracy configurable adder designs can be subtle and therefore is worth some discussion.

1) **Area**: In the literature, the area sometimes refers to the part of the circuit working in a certain mode, e.g., the circuit for the accurate part is not included in area estimation when evaluating approximation mode. We reporttheroutedlayoutareaofeachentiredesign.
2) **Delay**: Some configurable adders, such as ACA [15] and GeAr [16], implement error correction with pipelining, which sometimes takes multiple clock cycles to determine the complete result. The delay or performance evaluation of such designs is much more complicated than unpipelined designs. Our work is focused on unpipelined implementation, although it can be pipelined. Thus, the reported delay is the maximum combinational logic path delay obtained from Synopsys PrimeTime with consideration of wiredelay.
3) **Power**: The power dissipation is estimated by Synopsys PrimeTimeconsideringbothstaticanddynamicpower.
4) **Accuracy**: We use PSNR (Peak Signal-to-Noise Ratio), where errors are treated as noise, as a composite accuracy metric for considering both error magnitude and error rate. In addition, the worst case error, which is equivalent to the maximum error magnitude [13], and error rate are also reported. Each error result is from 100K-run Matlab-based Monte Carlo simulation assuming uniform distribution ofaddends.
5) **Tunability**: This means the range and granularity of runtime accuracy configurations. Sometimes, this can be confused with design-timeflexibility.
6) **Tradeoff**: The tradeoff among the above factors is complex and is difficult to capture in a simple picture. To this end, we use composite metrics including power-delay product (PDP), energy-delay product (EDP) and iso-delaypower.

B. Results of Tradeoff for DifferentConfigurations

In this part, we mainly compare the following accuracy configurable adder designs:

- GDA [18]: We use the same design as in [18], whereeachsub-adderhas4bits.Thisdesigncanbeconfigured by choosing accurate or predicted carry-out for each sub-adder. The carry prediction at each segment can also be configured to different accuracy levels by using different number of lower-bit addends.
- RAP-CLA [19]: We implement four different designs with carry prediction bit-width from 1 bit to 4 bits, which is reflected in the name. For example, RAP-CLA2 means each of its carry prediction is from its 2 lower bits. As in [19], each design can be configured to either only one approximation mode or accuratemode.
- SARA: This is our proposed design and we evaluate sub-adder bit-width of 1 bit, 4 bits and 8 bits, referred to as SARA1, SARA4 and SARA8,respectively.

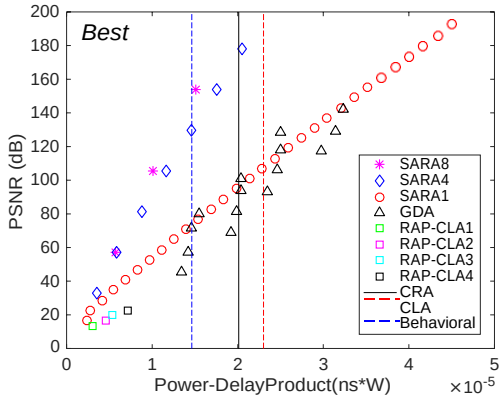


Fig. 8. SARA: PSNR versus power-delay product.

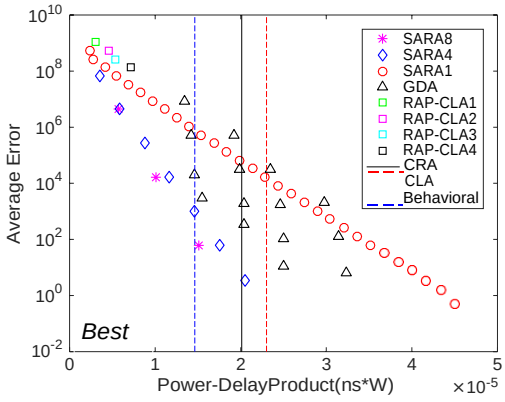


Fig. 9. SARA: Average error versus power-delay product.

The main result is shown in Figure 8, where each point is from one configuration of one design. The computation accuracy is evaluated by PSNR while the conventional design objectives are characterized by PDP. A design and configuration is ideal if it has large PSNR but low PDP, i.e., northwest in the figure. PDPs of two classic accurate designs, CRA and CLA, are indicated by the two vertical lines as their PSNR is near infinity. The result of SARA working in completely accurate mode is unable to be presented in the figure, because its infinite PSNR cannot be displayed as a single dot in theplot.Evidently,thebestsolutionsarefromSARA4and

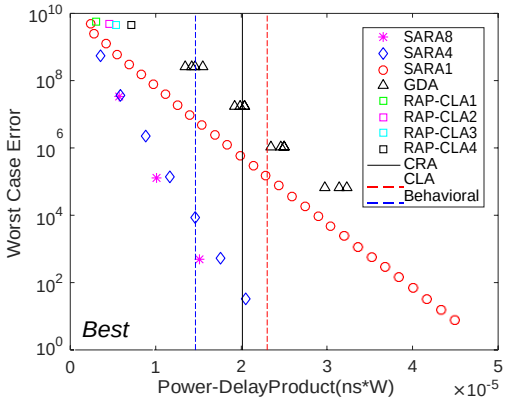


Fig.10. SARA:Theworstcaseerrorversuspower-delayproduct.

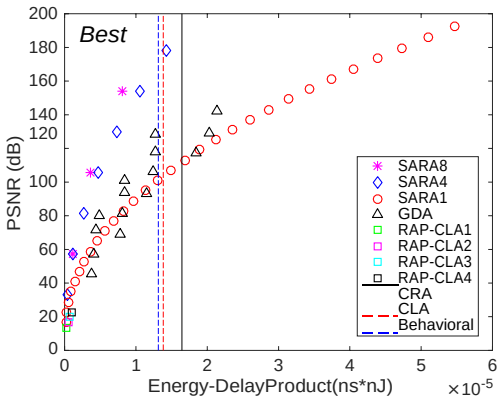


Fig. 11. SARA: PSNR versus energy-delayproduct.

SARA8. At 100*dB* PSNR, the PDP of SARA4 and SARA8 is about a half of GDA or CRA. The solutions from RAP-CLA, the latest previous work, are also largely dominated by SARA in PSNR-PDP tradeoff. An interesting case is SARA1. Its tradeoff is similar as GDA and not as good as SARA4 or SARA8. However, its runtime tunability is superior to all the other designs. It has the largest tuning range, the finest tuning granularity and very smooth tradeoff.

Figure 9 and 10 show the tradeoff between error magnitude and power-delay product. Ideally a better design or configuration has smaller average error or worst case error with lower PDP, which can be marked in the lower left corner of thefigure.InFigure9,SARA4andSARA8dominateother designs in average error-PDP tradeoff. For each configuration, SARA4 and SARA8 have almost the lowest average error at a certain PDP level. Although SARA1 cannot achieve superior average error and PDP tradeoff to GDA, it shows fine-grant tunability in a large range same as PSNR-PDP tradeoff. Figures 10 depicts the worst case error versus PDP and confirms the trend observed in the PSNR-PDP tradeoff. All SARA designs even for SARA1 have lower worst case error than previous work at the same PDP level. In addition, the result of SARA working in accurate mode cannot be found in the plot. That's because the y-axis is in Logarithmic scale and zero error will be converted into infinite which cannot be displayed as a singledot.

EDP is another metric to efficiently evaluate tradeoffs betweencircuitlevelpowersavingtechniquesfordigitaldesigns. Figure 11 to 13 illustrate accuracy versus EDP, which have similar trend in accuracy-PDP tradeoff. Most configurations of SARA4 and SARA8 have lower EDP than accurate adder CRA and CLA. At a certain EDP level, SARA4 and SARA8 still dominate GDA and RAP-CLA with larger PSNR, smaller average error or worst case error. SARA1 in different configurations cover the range from lowest to highest EDP, which provides finest tunability in accuracy-energy tradeoff among differentarchitectures.

C. Results of Tradeoff for Delay-AdaptiveReconfiguration

This part is to evaluate the SARA-DAR design, where the configuration decision has already been made. Hence, it makes

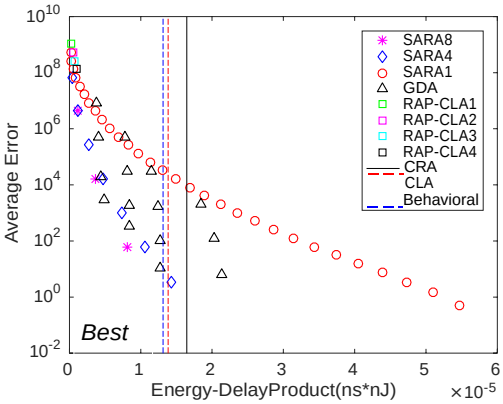


Fig. 12. SARA: Average error versus energy-delay product.

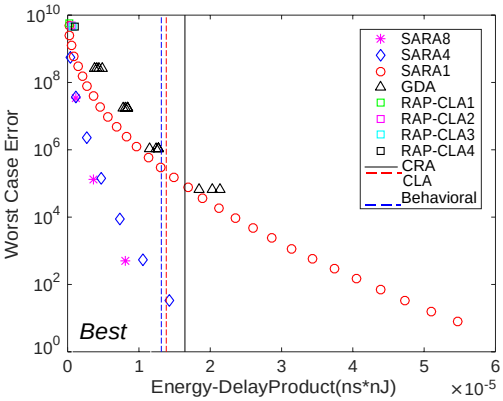


Fig. 13. SARA: The worst case error versus energy-delay product.

sense to additionally compare with static approximate adders, where no configuration is needed. Static approximate adder designs including ETAII [8], FICTS [13] and AFICTS [13] are implemented in the experiment. In addition, CRA-based approximate designs CRA-trunc*i* implemented by ignoring lowest *i* bits in addends are presented, which is a simple but good baseline for comparison. Seven SARA4-DAR designs are obtained based on seven configurations of SARA4 with detection window of 2 bits, while three SARA8-DARare

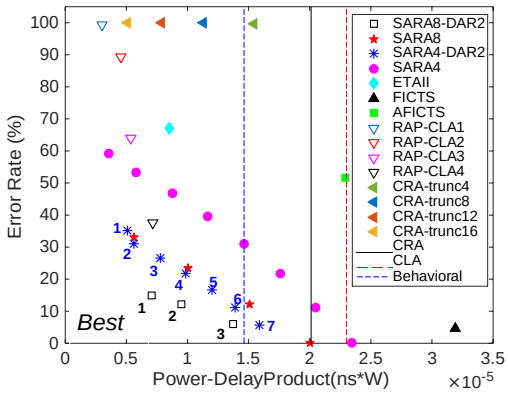


Fig. 14. SARA-DAR: Error rate versus power-delayproduct.

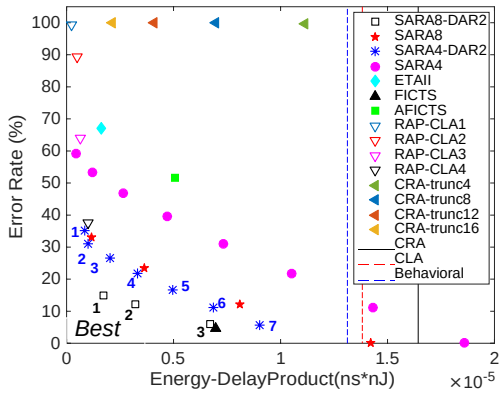


Fig.16.SARA-DAR:Errorrateversusenergy-delayproduct.

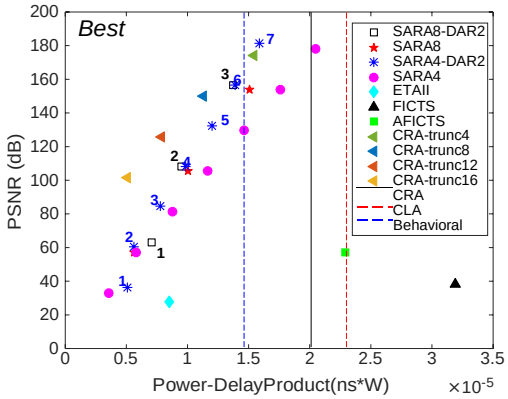


Fig. 15. SARA-DAR: PSNR versus power-delay product.

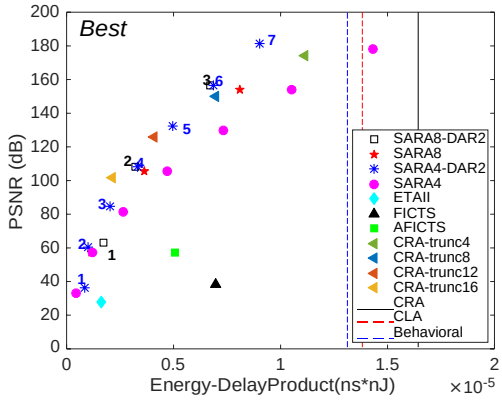


Fig. 17. SARA-DAR: PSNR versus energy-delay product.

basedondifferentconfigurationsofSARA8.Thatis,ifaMUX at bit *i* is configured to accurate carry in SARA4/SARA8, bit *i* of corresponding SARA-DAR is hard-wired to accurate carry without using MUX. When bit *j* in SARA4/SARA8 isin approximation mode, bit *j* of corresponding SARA-DAR uses the delay-adaptivereconfiguration.

Figure 14 shows the error rate versus PDP tradeoff. The dot of SARA4-DAR2 labeled with '1' represents the counterpart of SARA4 when all the MUXes are controlled by delay-adaptive reconfiguration. When we remove the MUX at the highest bit to propagate accurate carry and keep others in delay-adaptive reconfiguration, another SARA4-DAR2 design could be obtained (another dot labeled with '2' in the figure). If we go on to remove more MUXes in MSB, a series of SARA4-DAR2 designs shown as dots with label '3' to '7' can be obtained. Three SARA8-DAR2 designs are created in the same way. According to the figure, the error rate of SARA is mostly lower than RAP-CLA. By using delay-adaptive reconfiguration, SARA-DAR often has less error rate and PDP than SARA. SARA-DAR also greatly outperforms the static approximate adders in both error rate and PDP. Moreover, in Figure 14 we can observe those dots right on the x-axis which represent SARA4 and SARA8 working in accurate mode. Both of them achieve zero error rate. PDP ofaccurateSARA8isabout$2\times10^{-5}ns\cdot W$,whilePDPof accurate SARA4 is almost $2.3\times10^{-5}ns\ W$. In Figure 15, SARA-DARalsodemonstratesbetterPSNR-PDPtradeoffthan other designs, except for comparing with CRA-trunc at some low-PSNRlevels.However,CRA-trunchasalmost100%error rate since it dismiss lower bits in addends, which is the worst among all static approximate adders. Figure 16 and 17 show tradeoff between accuracy and EDP. At a certain EDP level, SARA-DAR has almost the same PSNR as CRA-trunc, which isthebestamongallstaticapproximateadders.

D. Impact of Detection Window in Delay-Adaptive Reconfiguration

Thispartshowstheimpactofdetectionwindowinthetrade- off for delay-adaptive reconfiguration. According to Equa- tion (23), the error rate of MUX output can be reduced by delay-adaptive reconfiguration. As the length of detection window increases, the error rate would decrease because there are less probability that MUX is configured in approximate mode. As a result, the overall error rate varies with the size of detection window. Figure 18 and 19 show the changes of error rate and PSNR of SARA4-DAR with different de-tection window. As the size of detection window increases from 1 to 3, the error rate decreases compared to its SARA counterpart. However, we can observe that the gap of error ratebetweenSARA4-DAR2andSARA4-DAR1varieswith

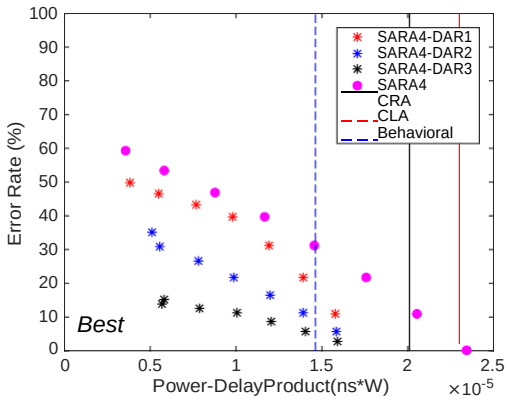


Fig.18. ErrorrateofSARA4-DARwithdifferentdetectionwindow.

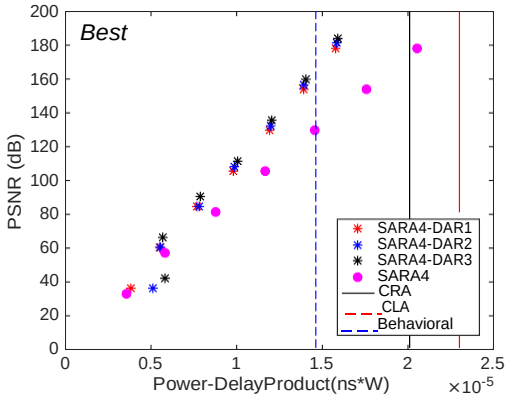


Fig. 19. PSNR of SARA4-DAR with different detection window.

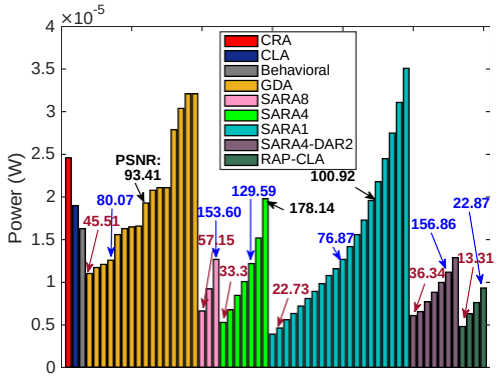


Fig.20.Iso-delaypowercomparison.ThenumbersarePSNR.

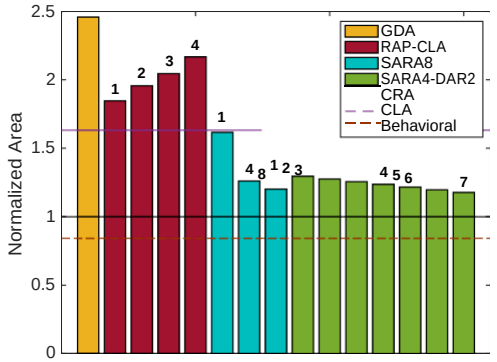


Fig. 21. Area comparison.

different configuration. Although the change in error rate for individual MUX of SARA-DAR is proportional to the size of detection window (as shown in Equation (23)), the overall error rate in output results might not show linear change. When the size of detection window increases by 1, PSNR of SARA4-DAR increases by about 3dB on average. We can also find that the PDP gap between SARA4-DAR2 and SARA4-DAR1 varies with different configurations in both figures. The change of PDP between SARA4-DAR2 and SARA4-DAR1 in most configurations is very small, while it's larger in the first configuration (which is presented as the first dot of SARA4-DAR in the left of the figures). It is mainly attributed to unproportioned change in delay between different configurations.

E. Results of Iso-delay Power and Area

Although power-delay product results have been shown in Sections VI-B and VI-C, the tradeoff between power anddelay is still unclear. The power-delay tradeoff can be obtained by different accuracy configurations or varying supply voltages. Different combinations of configurations and voltages may lead to overwhelming volume of results, which are difficult to interpret, especially when implication to accuracy is involved at the same time. Thus, we indicate the tradeoff by investigatingtheiso-delaypower,whichisthepowerofeach circuit tuned to the same critical path delay ($0.82ns$) by voltage scaling. The results are shown in Figure 20. Ingeneral, SARA4, SARA8 and SARA4-DAR can achieve much lower power than CRA. Although GDA and RAP-CLA seem to provide low power, their PSNR is much less than our designs. Compared at the same iso-delay power level, SARA has more than 20dB increase in PSNR than RAP-CLA, while GDA has more than 70dB decrease than SARA designs. SARA1 shows a large range of iso-power tuning which could reach the lowest and highest power among all adders. We do not have iso-delay power for approximate adders working in accurate mode, because the delay of such case is larger than CRA due to induction of MUXes which cannot provide sufficient room for reducing supplyvoltage.

Last but not the least, we compare area of these designs in Figure 21. Same as our expectation, GDA and RAP-CLA have greater area than CLA while area of SARA4 or SARA8 is significantly smaller than CLA. SARA1 has almost thesame areaasCLAduetoMUXesineverybitwhichaidtheaccuracy configuration. On average, the area of SARA is 39% smaller thanthatofRAP-CLAand50%smallerthanthatofGDA.

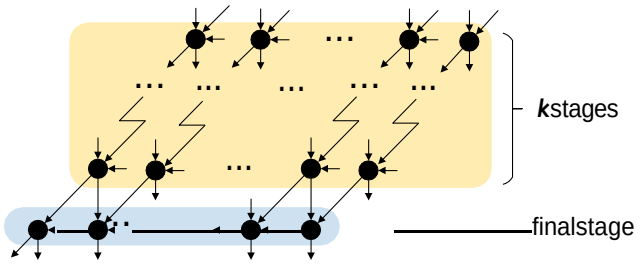


Fig. 22. Basic structure of multiplier.

VII. Applications

A. Extension toMultiplier

In complicated datapath system, multiplier is considered as a much bigger component in power consumption. Our carry-prediction-based approximation uses generate bit to predict the carry from lower segments. The critical delay can be restrainedtoasmallervaluewithshortercriticalpathincarry propagation. Further extension of our technique to multiplier depends on the multiplication structure used in hardware implementation. There is a variety of hardware designs for multiplication, according to the structures of reduction tree. In this section, we apply our technique on three kinds of multiplication structures including array multiplier, Wallace multiplier and Daddamultiplier.

As shown in Figure 22, the basic structure of multiplier employs a three-step process to multiply two integers.

- Step 1: Generate all partial products by using an AND gatearray.
- Step 2: Combine the partial products in *k* stages by layers ofhalf/fulladderuntilthematrixheightisreducedtotwo. Different types of structures depend on the reduction tree used to reduce the number of partial products in thisstep.
- Step 3: Sum the resulting numbers in the final stage by a conventionaladder.

In array multiplier the carry bits in one stage are propagated diagonally downwards, which follows the basic shift-and-add multiplication algorithm. Wallace multiplier based on Wallace tree combines the partial products as early as possible, which makes it faster than array multiplier [20]. Dadda's strategy is to make the combination take place as late as possible, which leads to simpler reduction tree and wider adder in final stage [20]. Thus, we can design approximate multipliers by using ourSARAdesigninsteadofCRAinthefinalstage.

Three types of 16×16 multipliers (array multiplier, Wallace multiplier and Dadda multiplier) as well as behavioral multiplier are synthesized and implemented by using Nangate 45nm Open Cell Library. Their error data are obtained from 100K-run Monte Carlo simulation with uniform distribution of operands. In approximate multiplier the final stage uses SARA4 which consists of sub-adders with bit-width of 4 bits, while the accurate one uses CRA. Figure 23 and 24 present the tradeoff between error and PDP. Most of approximate multipliers configured in approximate mode have better PDP comparedwiththeaccuratemultipliers.Thevarianceoferror between different approximate mode in approximatemultiplier has similar trend as SARA. Total error increases as more bits are configured in approximate mode. Approximate array multiplier shows larger error than approximate Wallace/Dadda multiplier at the same PDP level. It's because array multiplier has larger critical delay from internal stages in step 2 than Wallace/Daddamultiplier.

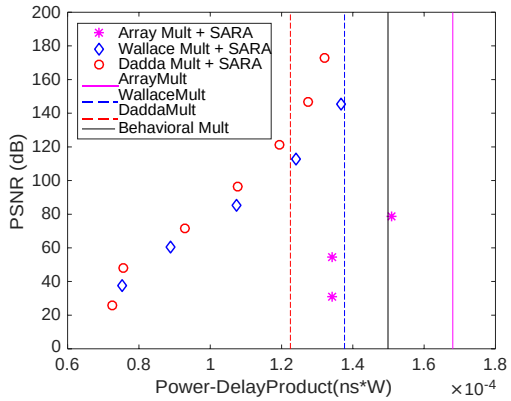


Fig. 23. Multiplier: PSNR versus power-delay product.

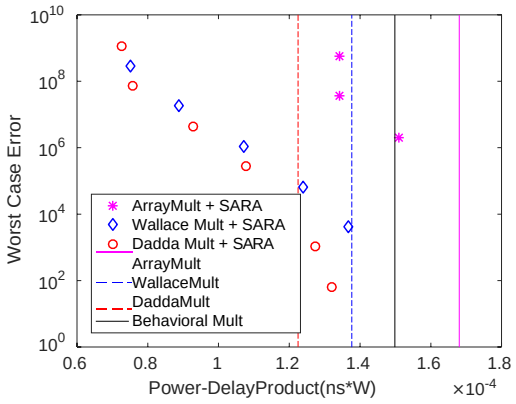


Fig.24. Multiplier:Theworstcaseerrorversuspower-delayproduct.

Figure 25 and 26 show the error versus EDP for both accurate and approximate multipliers. As more MUXes are set to propagate approximate carry, the average error in output increases to about 10^7, which as well achieves best EDP. The worst-case error rate of approximate Dadda multiplier is about 30%, while it comes to about 17% for approximate array multiplier and Wallace multiplier. As shown in Figure 26, when approximate multipliers are working in completely accurate mode (error rate equals 0), EDP is larger than that of their accurate counterpart. In summary, The experimental results show that our technique can be successfully extended to high speed multiplier designs. And due to the simple but effective structure of SARA it provides an easy way for us to convert conventional multiplier into approximatedesign.

B. DCT Computation in ImageProcessing

The discrete cosine transform (DCT) has been recognized asthebasicinmanytransformcodingmethodsforimageand

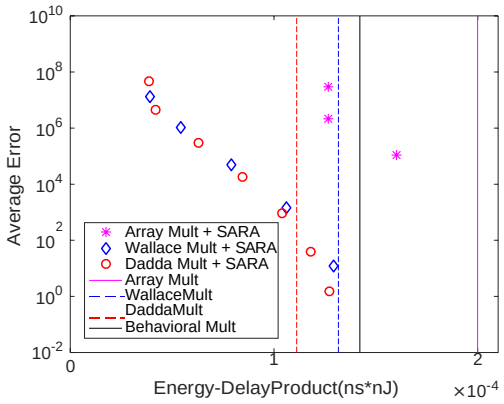


Fig. 25. Multiplier: Average error versus energy-delayproduct.

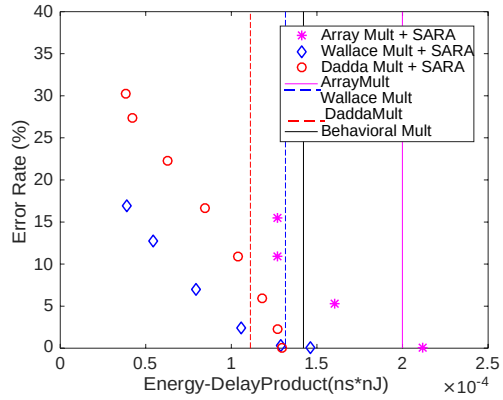


Fig. 26. Multiplier: Error rate versus energy-delay product.

video signal processing. It is used to transform the pixel data of image or video into corresponding coefficients in frequency domain. Since human visual system is more sensitive tothe changes in low frequency, the lost of accuracy in high-frequency components does not heavily degrade the quality of image processed by DCT. In addition, those components in different frequency have different tolerances to thedegradation inoriginaldata.Itisagoodexampletoshowthereconfigurability of our design by applying them in VLSI implementation of DCT computing in JPEG image compression.

The two-dimensional DCT is implemented by the row-column decomposition technique, which contains two stages of 1-D DCT [21]–[23]. The 2-D DCT of size $N \times N$ could be definedas

$$Z = C^t X C \tag{24}$$

where C is a normalized N th-order matrix and X is the data matrix. Generally the image is divided into several $N \times N$ blocks and each block is transformed by 2-D DCT into frequencydomaincomponents.TheVLSIimplementationof DCT computing contains a set of ROM and Accumulator Components(RACs)whichcanbeimplementedbymultipliers andadders[21]–[23].Inthisapplicationweuseapproximate adderstoreplacethoseaccurateonesinRCAstoimplement animprecisebutlowpowercircuitforimageprocessingwhich contains DCTcomputing.

TABLE IV
IMAGE QUALITY COMPARISON IN PSNR

	lenna	cameraman	kiel	house	AVERAGE
Accurate	39.85	38.23	37.68	37.35	38.27
SARA4	38.32	37.50	36.83	36.53	37.30
SARA8	35.33	35.07	34.92	34.81	35.03
SARA4-DAR2	39.45	37.90	37.43	37.00	37.97
GDA	34.53	34.55	34.88	34.20	34.54
RAP-CLA	33.38	33.44	33.51	33.39	33.43

We replace the adders in circuits with different configurations of SARA, SARA-DAR, GDA as well as RAP-CLA.The results are obtained by numerical simulation in Matlab. Aswe know, after DCT process data in different frequency domain have different level of error tolerance. As shown in Figure 31, matrix components in the upper-left corner correspond to lower frequency coefficients which are sensitive to human vision, while those components in lower-right corner might allow moreerrors.

To utilize this feature for better energy-accuracy tradeoff, wemakefollowingconfigurationfordifferentdesigns.

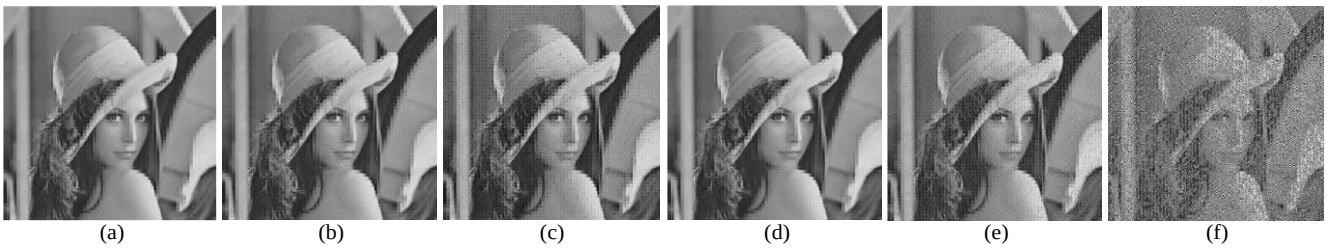
1) **SARA4**: SARA4 with 4, 3, 2, 1 consecutive segments working in accurate mode are used to compute components in S_1, S_2, $S_3$and S_4respectively.
2) **SARA8**: SARA8 with 1 segments in accurate mode are used to compute components in S_1, S_2, while another configuration with all segments in approximate mode are for S_3,S_4.
3) **SARA4-DAR2**: DAR counterpart of SARA4 with detection window of 2bits.
4) **GDA**: $GDA_{4,1}$, $GDA_{3,1}$, $GDA_{2,1}$, $GDA_{1,1}$(same notation as [18]) are used to compute components in S_1, S_2, $S_3$and S_4respectively.
5) **RAP-CLA**: since RAP-CLA can work in one approximate mode, we use RAP-CLA with window size of 20, 16,12,8tocomputecomponentsinS_1,S_2,S_3andS_4.

The image processing results are shown in Table IV. PSNR inthetableisdefinedviathemeansquarederror(MSE).Given an $m \times n$ image I and its restored image K, MSE and PSNR are definedas

$$MSE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} [I(i,j) - K(i,j)]^2 \tag{25}$$

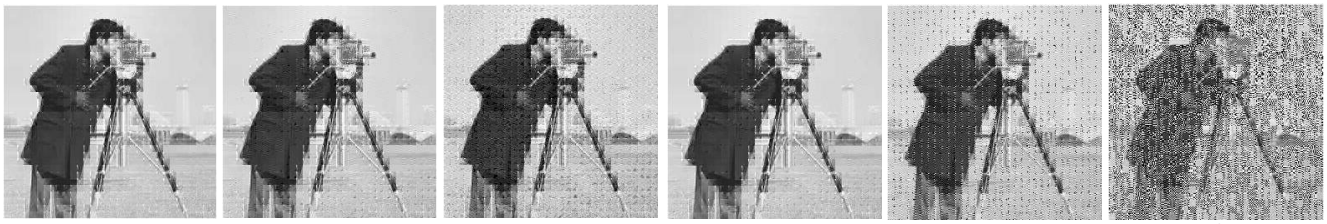
$$PSNR = 20 \cdot \log(MAX_I) - 10 \cdot \log(MSE), \tag{26}$$

where $MAX_I$is the maximum pixel value of the image. SARA4-DAR2 has the highest PSNR for every image among all configurable adders, which is close to the quality of accurate adder. Comparing SARA8 with GDA, they have similar PSNR and similar delay, but SARA8 has less power consumption according to the analysis in the previous section. SARA4-DAR2 achieves better image quality than SARA4, but might result in more power due to additional logics for self-configuration. The image quality for different adders in DCT computing can also be demonstrated in Figure 27 to 30. According to human vision, SARA and its DAR counterpart show better image quality than GDA and RAP-CLA in JPEG compression processing.



(a) (b) (c) (d) (e) (f)

Fig. 27. Comparison of image lenna: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.

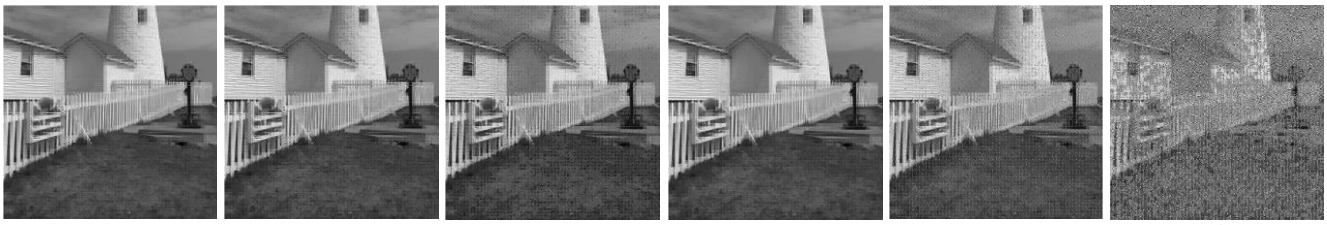


(a) (b) (c) (d) (e) (f)

Fig. 28. Comparison of image cameraman: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.

(a) (b) (c) (d) (e) (f)

Fig. 29. Comparison of image kiel: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.



(a) (b) (c) (d) (e) (f)

Fig. 30. Comparison of image house: (a) accurate adder; (b) SARA4; (c) SARA8; (d) SARA4-DAR2; (e) GDA; (f) RAP-CLA.

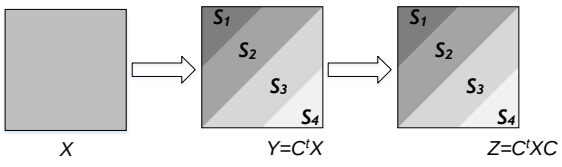


Fig. 31. 2 dimensional descrete cosine transform.

VIII. Conclusion

In this paper, we propse a simple accuracy reconfigurable adder(SARA)design.Ithassignificantlylowerpower/energy-delay product than the latest previous work when comparingat the same accuracy level. In addition, SARA has considerable lower area overhead than almost all the previous works. The accuracy-power-delay efficiency is further improved by a delay-adaptive reconfiguration technique. We demonstrate the efficiency of our adder in the applications of multiplication circuitsandDCTcomputingcircuitsforimageprocessing.

Acknowledgment

The authors would like to thank Dr. Duncan M. Walker from Texas A&M University for helpful discussions.

References

[1]J. Han and M. Orshansky, “Approximate computing: An emerging paradigm for energy-efficient design,” in *IEEE European Test Symposium*, pp. 1–6, 2013.

[2]V. Gupta, D. Mohapatra, A. Raghunathan, and K. Roy, "Low-power digital signal processing using approximate adders," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 32, no. 1, pp. 124–137,2013.
[3]V. Chippa, A. Raghunathan, K. Roy, and S. Chakradhar, "Dynamic effort scaling: Managing the quality-efficiency tradeoff," in *Design Automation Conference (DAC)*, pp. 603–608, 2011.
[4]S.-L. Lu, "Speeding up processing with approximation circuits," *Computer*,vol.37,no.3,pp.67–73,2004.
[5]S. Venkataramani, K. Roy, and A. Raghunathan, "Substitute-and-simplify: A unified design paradigm for approximate and quality configurable circuits," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1367–1372,2013.
[6]K. Du, P. Varman, and K. Mohanram, "High performance reliable variable latency carry select addition," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1257–1262, 2012.
[7]A. K. Verma, P. Brisk, and P. Ienne, "Variable latency speculative addition: A new paradigm for arithmetic circuit design," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1250–1255, 2008.
[8]N. Zhu, W. L. Goh, and K. S. Yeo, "An enhanced low-power high-speed adder for error-tolerant application," in *International Symposium onIntegratedCircuits*,pp.69–72,2009.
[9]N. Zhu, W. L. Goh, W. Zhang, K. S. Yeo, and Z. H. Kong, "Design of low-power high-speed truncation-error-tolerant adder and its application in digital signal processing," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 8, pp. 1225–1229, 2010.
[10]N. Zhu, W. L. Goh, and K. S. Yeo, "Ultra low-power high-speed flexible probabilistic adder for error-tolerant applications," in *International SoC Design Conference (ISOCC)*, pp. 393–396, 2011.
[11]G. Liu, Y. Tao, M. Tan, and Z. Zhang, "Casa: correlation-aware speculative adders," in *International Symposium on Low Power Electronics and Design*, pp. 189–194, 2014.
[12]J. Hu and W. Qian, "A new approximate adder with low relative error and correct sign calculation," in *Conference on Design, Automation and Test in Europe (DATE)*, pp. 1449–1454,2015.
[13]J. Miao, K. He, A. Gerstlauer, and M. Orshansky, "Modeling and synthesis of quality-energy optimal approximate adders," in *International Conference on Computer-Aided Design (ICCAD)*, pp. 728–735, 2012.
[14]S. Mazahir, O. Hasan, R. Hafiz, M. Shafique, and J. Henkel, "An area-efficient consolidated configurable error correction for approximate hardware accelerators," in *Design Automation Conference (DAC)*, pp. 1–6, 2016.
[15]A. B. Kahng and S. Kang, "Accuracy-configurable adder for approximate arithmetic designs," in *Design Automation Conference (DAC)*, pp. 820–825, 2012.
[16]M. Shafique, W. Ahmad, R. Hafiz, and J. Henkel, "A low latency generic accuracy configurable adder," in *Design Automation Conference (DAC)*, pp. 1–6, 2015.
[17]V. Benara and S. Purini, "Accurus: A fast convergence technique for accuracy configurable approximate adder circuits," in *IEEE Computer Society Annual Symposium on VLSI*, pp. 577–582, 2016.
[18]R. Ye, T. Wang, F. Yuan, R. Kumar, and Q. Xu, "On reconfiguration-oriented approximate adder design and its application," in *International Conference on Computer-Aided Design (ICCAD)*, pp. 48–54, 2013.
[19]O. Akbari, M. Kamal, A. Afzali-Kusha, and M. Pedram, "RAP-CLA: A reconfigurable approximate carry look-ahead adder," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. PP, no. 99, pp. 1–1, 2017.
[20]W. J. Townsend, E. E. Swartzlander, and J. A. Abraham, "A comparison of dadda and wallace multiplier delays," *Proc. SPIE, Advanced Signal Processing Algorithms, Architectures, and Implementations*, vol. 13, pp. 552–560,2003.
[21]S. Yu and E. Swartziander, "DCT implementation with distributed arithmetic," *IEEE Transactions on Computers*, vol. 50, no. 9, pp. 985–991, 2001.
[22]M.-T. Sun, T.-C. Chen, and A. M. Gottlieb, "VLSI implementation of a 16* 16 discrete cosine transform," *IEEE Transactions on Circuits and Cystems*, vol. 36, no. 4, pp. 610–617, 1989.
[23]D. Gong, Y. He, and Z. Cao, "New cost-effective VLSI implementation of a 2-d discrete cosine transform and its inverse," *IEEE Transactions on Circuits and Systems for VideoTechnology*, vol. 14, no. 4, pp. 405–415.

Wenbin Xu received the B.S. and M.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2008 and 2011 respectively. He is now working toward the Ph.D. degree in computer engineering at Texas A&M University, College Station, TX. He is currently a Research Assistant in the Department of Electrical and Computer Engineering, Texas A&M University. His research interests include computer-aided design, approximate computing and hardwaresecurity.

Sachin Sapatnekar (S'86, M'93, F'03) receivedthe B. Tech. degree from the Indian Institute of Technology, Bombay, the M.S. degree from Syracuse University, and the Ph.D. degree from the University of Illinois. He taught at Iowa State University from 1992 to 1997 and has been at the University of Minnesota since 1997, where he holds the Distinguished McKnight University Professorship and the Robert and Marjorie Henle Chair. He has received seven conference Best Paper awards, a Best Poster Award, two ICCAD 10-year Retrospective Most Influential Paper Awards, the SRC Technical Excellence award and the SIA University Researcher Award. He is a Fellow of the ACM and the IEEE.

Jiang Hu (F'16) received the B.S. degree in optical engineering from Zhejiang University (China) in 1990, the M.S. degree in physics in 1997 and the Ph.D. degree in electrical engineering from the University of Minnesota in 2001. He worked with IBM Microelectronics from January 2001 to June 2002. In 2002, he joined the electrical engineering faculty at Texas A&M University. His research interest is in optimization for large scale computing systems, especially VLSI circuit optimization, adaptive design, hardware security, resource allocation andpower management in computing systems.

Dr.HureceivedtheBestPaperAwardattheACM/IEEEDesignAutomation Conference in 2001, an IBM Invention Achievement Award in 2003, and the Best Paper Award from the IEEE/ACM International Conference on Computer-Aided Design in 2011. He has served as a technical program committee member for DAC, ICCAD, ISPD, ISQED, ICCD, DATE, ISCAS, ASP-DAC and ISLPED. He is general chair for the 2012 ACM International Symposium on Physical Design, and an associate editor of IEEE transactionson CAD 2016-2011. Currently he is an associate editor for the ACM transactions on Design Automation of Electronic Systems. He received a Humboldt research fellowship in2012.

Carbon Nanotube Using Electronics

1Kukatla Rajeshwar 2Singamaneni Nagajyothi

Abstract

We evaluate the potential of carbon nanotubes (CNTs) as the basis for a new nanoelectronic technology. After briefly reviewing the electronic structure and transport properties of CNTs, we discuss the fabrication of CNT field-effect transistors (CNTFETs) formed from individual single-walled nanotubes (SWCNTs), SWCNT bundles, or multiwalled (MW) CNTs. The performance characteristics of the CNTFETs are discussed and compared to those of corresponding silicon devices. We show that CNTFETs are very competitive with state-of-the-art conventional devices. We also discuss the switching mechanism of CNTFETs and show that it involves the modulation by the gate field of Schottky barriers at the metal-CNT junctions. This switching mechanism can account for the observed subthreshold and vertical scaling behavior of CNTFETs, as well as their sensitivity to atmospheric oxygen. The potential for integration of CNT devices is demonstrated by fabricating a logic gate along a single nanotube molecule. Finally, we discuss our efforts to grow CNTs locally and selectively, and a method is presented for growing oriented SWCNTs without the involvement of a metal catalyst.

Keywords*—*Carbon nanotubes (CNTs), field-effect transistors (FETs), molecular electronics, nanoelectronics.

I. INTRODUCTION

Carbon nanotubes (CNTs) are hollow cylinders composed of one or more concentric layers of carbon atoms in a honeycomb lattice arrangement. Multiwalled nanotubes (MWCNTs) were observed for the first time in transmission electron microscopy (TEM) studies by Iijima in 1991 [1], while single-walled nanotubes (SWCNTs) were produced independently by Iijima [2] and Bethune [3] in 1993.

SWCNTs typically have a diameter of 1–2 nm and a length of several micrometers. The large aspect ratio makes the nanotubes nearly ideal one-dimensional (1-D) objects, and as such the SWCNTs are expected to have all the unique properties predicted for these low-dimensional structures [4]–[7]. In addition, as we discuss below, depending on the detailed arrangement of the carbon atoms the SWCNTs can be metallic or semiconducting [8], [9]. Furthermore, the C–C bonds in CNTs are very strong, resulting in an extremely high mechanical stability (Young's modulus about ten times higher than that of steel) and chemical inertness. The strong, covalent bonding also leads to near perfect side-wall structures with very few defects.

CNTs are currently considered as promising building blocks of a future nanoelectronic technology. This is not simply due to their small size but rather to their overall properties. In fact, many of the problems that silicon technology is or will be facing are not present in CNTs. Below we list some of these CNT properties and their implications for electronics.

1) Carrier transport is 1-D. This implies a reduced phase space for scattering of the carriers and opens up the possibility of ballistic transport. Correspondingly, power dissipation is low. Furthermore, as we discuss in Section II, their electrostatic behavior is different from that of silicon devices with implications on screening and electron/hole tunneling.
2) All chemical bonds of the C atoms are satisfied and there is no need for chemical passivation of dangling bonds as in silicon. This implies that CNT electronics would not be bound to use SiO_2 as an insulator. High dielectric constant and crystalline insulators can be used, allowing, among other things, the fabrication of three-dimensional (3-D) structures.
3) The strong covalent bonding gives the CNTs high mechanical and thermal stability and resistance to electromigration. Current densities as high as 10^9 A/cm^2 can be sustained [10].
4) Their key dimension, their diameter, is controlled by chemistry, not conventional fabrication.
5) In principle, both active devices (transistors) and interconnects can be made out of semiconducting and metallic nanotubes, respectively.

We see that the properties of the SWCNTs are truly remarkable. However, finding ways to effectively exploit these properties remains a challenge. In the rest of this paper

Manuscript received October 2, 2002.

The authors are with the IBM Research Division, T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: avouris@us.ibm.com).

Digital Object Identifier 10.1109/JPROC.2003.818338

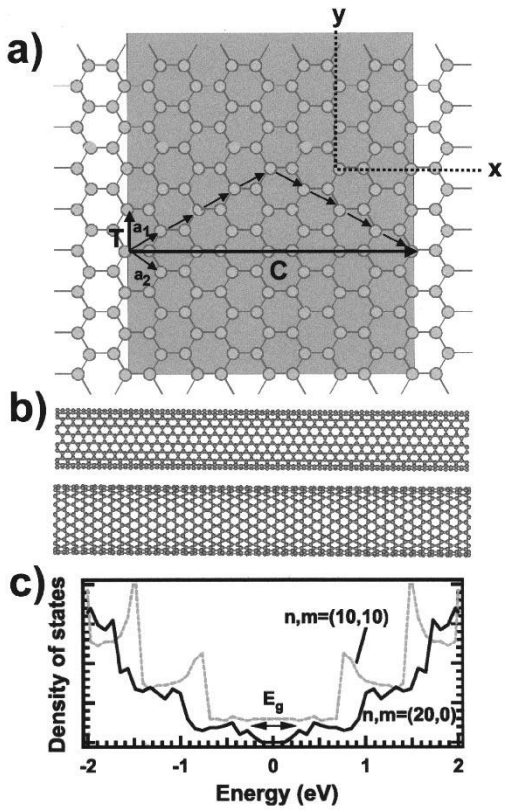


Fig. 1. (a) Description of the structure of CNTs in terms of the chirality vector C and the (n,m) indices. The example shown involves a (4,4) tube. (b) Atomic structure of a metallic (10,10) CNT (top) and a semiconducting (20,0) CNT (bottom), and (c) the corresponding density of states versus energy plots.

we will review the electronic structure of the CNTs and then present our efforts to fabricate nanotube transistors and simple integrated circuits and understand the underlying device physics.

II. Electronic Structure of Carbon Nanotubes

The electronic structure and electrical properties of SWCNTs are usually discussed in terms of the electronic structure of a graphene sheet (a layer of graphite) [4]–[9]. The SWCNT can be thought of as being formed by folding a piece of graphene to give a seamless cylinder. The circumference of the nanotube is expressed by the so-called chirality vector, $\mathbf{C}$, connecting two crystallographically equivalent sites of the two-dimensional (2-D) graphene sheet (see Fig. 1). $\mathbf{C} = n\mathbf{a}_1 + m\mathbf{a}_2$, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the unit vectors of the hexagonal honeycomb lattice, so that any nanotube can be described by a pair of integers (n,m) that define its chiral vector. For example, the chiral vector shown in Fig. 1(a) describes a (4,4) nanotube. The unit shell of the nanotube is defined as the rectangle formed by $\mathbf{C}$ and the 1-D translational vector $\mathbf{T}$ identified in Fig. 1. Also shown are the atomic structures of a (10,10) and a (20,0) CNT [Fig. 1(b)] and their density of states [Fig. 1(c)].

The interesting electrical properties of CNTs are due in a large part to the peculiar electronic structure of the graphene. Its band structure and its hexagonal first Brillouin zone

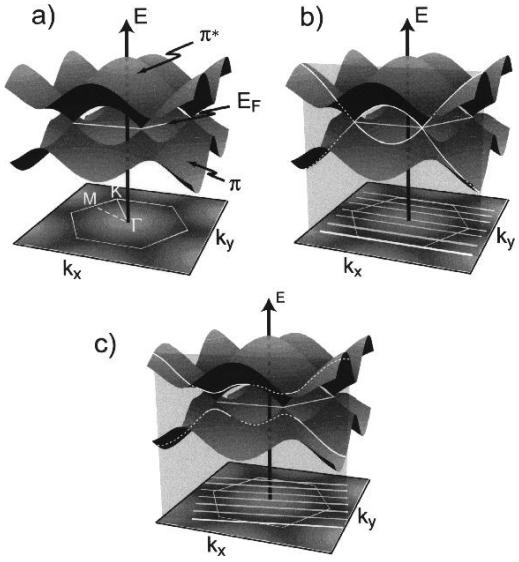


Fig. 2. (a). Band structure of a graphene sheet (top) and the first Brillouin zone (bottom). (b) Band structure of a metallic (3,3) CNT. (c) Band structure of a (4,2) semiconducting CNT. The allowed states in the nanotubes are cuts of the graphene bands indicated by the white lines. If the cut passes through a K point, the CNT is metallic; otherwise, the CNT is semiconducting.

are shown in Fig. 2(a). The energy surfaces describing the valence (π) and conduction (π^*) states touch at six points (Fermi points) lying at the Fermi level. This unusual band structure has immediate consequences for the electronic properties of graphene. While allowed states exist at the Fermi level, the dimensionality of the system (2-D) results in a vanishing density of states when integrating over the Fermi surface. Because of this particular situation, graphene is a zero-gap semiconductor.

In the case of a nanotube there is an additional quantization arising from the confinement of the electrons in the circumferential direction in the tube. This requires that the circumferential component of the wave vector $\boldsymbol{k}_C$ can only take the values fulfilling the condition $\boldsymbol{k}_C \cdot \mathbf{C} = 2\pi j$ where $\mathbf{C}$ is again the chirality vector and j an integer. As a result, each band of graphene splits into a number of 1-D subbands labeled by j. Fig. 2(b) shows the states of a (3,3) CNT. The allowed energy states of the tube are cuts of the graphene band structure. When these cuts pass through a Fermi point, as in the case of the (3,3) nanotube, the tube is metallic. In cases where no cut passes through a K point, the tubes are semiconducting [Fig. 2(c)]. It can be shown that an (n,m) CNT is metallic when $n = m$, it has a small gap (due to curvature-induced $\sigma - \pi$ mixing) when $n - m = 3i$ [11], where i is an integer, while CNTs with $n - m \neq 3i$ are truly semiconducting [8], [9].

As long as we restrict our interest to low energies (i.e., a few hundred meV from the Fermi energy E_F) the band structure of a metallic nanotube can be approximated by two sets of bands with a linear dispersion intersecting at k_F and $-k_F$ [see white lines in Fig. 2(b)]. Electrons with $dE/dk > 0$ move to the right, while electrons with $dE/dk < 0$ move to the left. In semiconducting CNTs the two bands do not cross at E_F, but a diameter-dependent band gap develops

with $E_{GAP} = (4\hbar v_F/3d_{CNT})$, where d_{CNT} is the tube's diameter and v_F the Fermi velocity [8], [9]. The above theoretical predictions have been confirmed experimentally by scanning tunneling spectroscopy [12], [13].

The SWCNTs are 1-D objects and as such their two-terminal conductance is given by Landauer's equation: [14], [15] $G = (2e^2/h) \sum_i^N T_i$ where $2e^2/h$ is the quantum of conductance and T_i is the transmission of a contributing conduction channel (subband). The sum involves all contributing conduction channels, i.e., channels whose energy lies between the electrochemical potentials of the left and right reservoirs to which the nanotube is connected. In the absence of any scattering, i.e., when all $T_i = 1$, the resistance $(R = 1/G)$ of a metallic SWCNT is $h/(4e^2) \approx 6.5$ kΩ because, as we discussed above, $N = 2$. This quantum mechanical resistance is a contact resistance arising from the mismatch of the number of conduction channels in the CNT and the macroscopic metal leads.

There is strong evidence that $T_i \approx 1$ in the case of metallic SWCNTs, so that these tubes behave as ballistic conductors [16]–[19]. This arises from the 1-D confinement of the electrons which allows motion in only two directions. This constraint along with the requirements for energy and momentum conservation severely reduces the phase space for scattering processes. However, in addition to the quantum mechanical contact resistance, there are other sources of contact resistance, such as those produced by the existence of metal-nanotube interface barriers, or poor coupling between the CNT and the leads. These types of resistance are very important and can dominate electrical transport in nanotubes, especially at low temperatures where typically they lead to charging and the observation of Coulomb blockade phenomena. Localization can also be induced by contacts to metallic electrodes, a fact that makes four-probe measurements very difficult, requiring special arrangements [19].

Unlike SWCNTs, the electrical properties of MWCNTs have received less attention. This is due to their complex structure; every carbon shell can have different electronic character and chirality, and the presence of shell–shell interactions [20], [21]. However, at low bias and temperatures, and when MWCNTs are side-bonded to metallic electrodes, transport is dominated by outer-shell conduction [10], [22], [23]. MWCNTs show 1-D or 2-D characteristics, depending on their diameter and the property considered.

III. Fabrication and Performance of Carbon Nanotube FETs

FETs, particularly in CMOS form, have been proven to be the most technologically useful device structures. It is, thus, natural that we have chosen to build such devices using CNTs. The first such devices were fabricated in 1998 [24], [25]. In these a single SWCNT was used to bridge two noble metal electrodes prefabricated by lithography on an oxidized silicon wafer as shown in Fig. 3. The SWCNT played the role of the "channel," while the two metal electrodes functioned as the "source" and "drain" electrodes. The heavily doped silicon wafer itself was used as the "gate" (back-gate). These CNTFETs behaved as p-type FETs (we will return to this point in Section V) and had an I(on)/I(off) current ratio of $\approx 10^5$. While functional, the devices had a high parasitic contact resistance (≥ 1MΩ), low drive currents (a few nanoamperes), low transconductance $g_m \sim 1$ nS, and high inverse subthreshold slopes S ~1–2 V/decade. To a large extent the unsatisfactory characteristics were due to bad contacts. The CNT was simply laid on the gold electrodes and was held by weak van der Waals forces. To improve the contacts we adapted a different fabrication scheme where the semiconducting SWCNTs (s-SWCNTs) were dispersed on an oxidized Si wafer, and the source and drain electrodes, now made of metals that are compatible with silicon technology such as Ti or Co, were fabricated on top of them [26]. Thermal annealing of the contacts, which in the case of Ti electrodes led to the formation of TiC, produced a stronger coupling between the metal and the nanotube and a reduction of the contact resistance [26], [27].

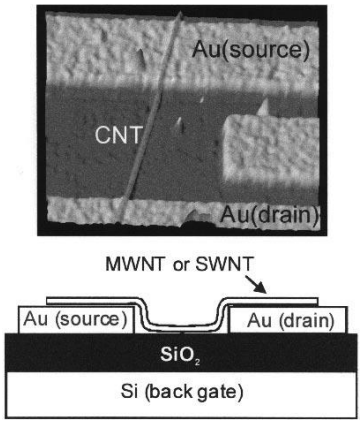


Fig. 3. Top: AFM image of one of our early CNTFETs. Bottom: Schematic cross section of the CNTFET [25].

Fig. 4(a) and (b) shows the output and transfer characteristics of such a CNTFET with Co electrodes [26]. From the $I_d - V_{gs}$ curves we see that the transistor is p-type and has a high on–off current ratio of $\sim 10^6$. This new CNTFET configuration has a significantly reduced contact resistance, ~ 30 kΩ, a much higher current in the μA range, and a transconductance $g_m = 0.34\mu$S, i.e., ~ 200 times higher than that of van der Waals-bonded CNTFETs.

All of the early devices were back-gated with very thick gate insulators (SiO_2 thickness $t_{ox} \approx$ 100-150 nm). As with conventional MOSFETs we should be able to improve their performance by increasing the gate capacitance by reducing the insulator thickness or increasing the dielectric constant. However, unlike in the MOSFET configuration where the capacitance is similar to that of a plane capacitor, i.e., gate capacitance $C_G \sim 1/t_{ox}$, the CNTFET geometry will predict a $C_G \sim \ln^{-1}(at_{ox} + b)$ dependence [25]. In addition to increasing the gate capacitance, it is essential that each CNTFET is gated independently by its own gate so that complex integrated circuits can be built.

A next generation of CNTFETs with top gates was fabricated by dispersing SWCNTs on an oxidized wafer. Atomic

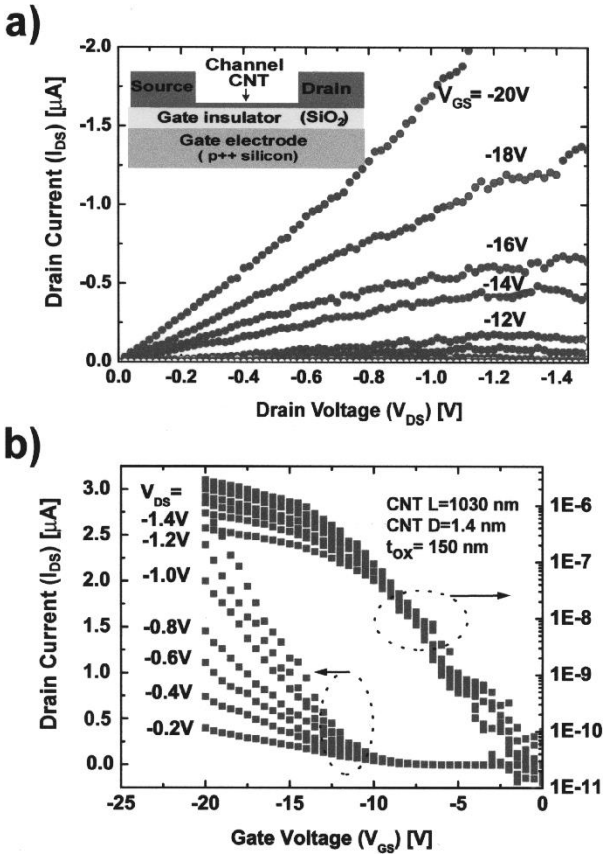


Fig. 4. (a) Output characteristics of a CNTFET with cobalt source and drain electrodes deposited on top of tube. (b) The corresponding transfer characteristics.

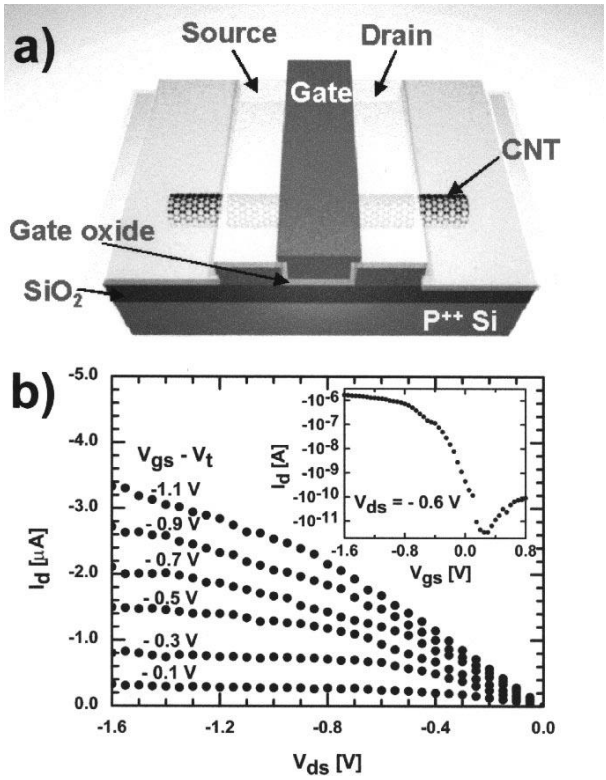


Fig. 5. (a) Schematic representation of one of our top-gated CNTFET with Ti source, drain, and gate electrodes. A 15-nm SiO_2 film was used as the gate oxide. (b) The *I–V* characteristics of the device.

force microscopy (AFM) imaging was used to identify single CNTs, and the Ti source and drain electrodes were fabricated on top of by e-beam lithography and liftoff [28]. After annealing at 850 °C to transform the contacts into TiC [27], a 15- to 20-nm-gate dielectric film was deposited by chemical vapor deposition (CVD) from a mixture of SiH_4 and O_2 at 300 °C. After annealing for ~0.5 h at 600 °C in N to $_2$ densify the oxide, 50-nm-thick Ti or Al gate electrodes were patterned by lithography and liftoff.

In Fig. 5(a) we show a schematic of a top-gated CNTFET, and in Fig. 5(b) the output characteristics of such a device with Ti electrodes and a 15 nm SiO_2 gate insulator film [28]. Such a CNTFET can also be switched by the bottom gate (wafer) and the resulting characteristics can be compared with those of the device under top-gate operation. This device has a superior performance; the threshold voltage of the top-gated CNTFET is significantly lower, –0.5 V, than under bottom-gated operation, +2 V. Similarly, the drive current is much higher under top gating, and the transconductance is similarly high, $g_m = 3.3$ μS per nanotube.

Since the eventual objective of nanotube electronics is to be competitive with silicon electronics, it is important to compare their relative performances, despite the fact that the CNTFETs are still far from being optimized. In these experiments a single SWCNT is used, so we express the current carrying capabilities of the devices per unit width (per micrometer) as is the practice in microelectronics. The

Table 1
Comparison of Key Performance Parameters for a 260 nm-Long Top Gate p-CNTFET, a 15-nm-Bulk Si p-MOSFET, and a 50- nm SOI p-MOSFET

	p-type CNTFET	Ref a	Ref. b
Gate Length (nm)	260	15	50
Gate oxide thickness (nm)	15	1.4	1.5
V_t (V)	-0.5	~-0.1	~-0.2
I_{ON} (μA/μm) ($V_{DS} = V_{GS} - V_t \sim -1$ V)	2,100	265	650
I_{OFF} (nA/μm)	-150	<-500	-9
Subthreshold slope (mV/dec)	130	~100	70
Transconductance (μS/μm)	2,321	975	650

a) B. Yu et al. IEDM Tech. Dig. 2001, p. 934; b) R. Chau et al. IEDM Tech. Dig. 2001, p. 621

diameter of the s-SWCNTs used was ~1.4 nm. Table 1 shows the comparison of the characteristics of the CNTFET with two recent high-performance Si p-channel devices: a 15-nm-gate length MOSFET built on bulk silicon [29] and a 50-nm-gate length device built using SOI technology [30]. We also note that an optimal device layout may require the use of an array of CNTFETs. In this case the results can be scaled to give the total current for the array [26]. The current would increase the denser the packing of the tubes. However, screening at close separations can reduce the actual current per CNT by up to a factor of 2 [31].

From Table 1 we see that the CNTFET is capable of delivering three to four times higher drive currents than the Si MOSFETs at an overdrive of 1 V, and has about four times higher transconductance. From the above and other considerations, it is clear that CNTFETs, even in this early stage of de-

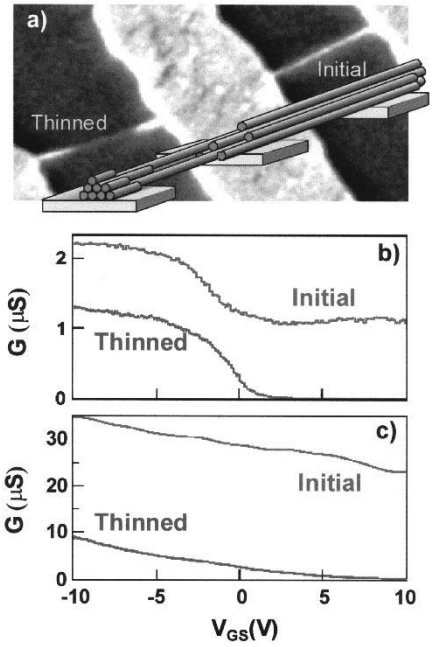


Fig. 6. Conductance (G) as a function of gate voltage (v_{GS}) of a CNT bundle containing both metallic and semiconducting nanotubes before and after selective breakdown of the metallic CNTs. (a) Images of the intact and thinned nanotube bundle. (b) G versus v_{GS} for a thin bundle. (c) The same for a very thick bundle. In the latter case some semiconducting tubes had to be sacrificed in order to remove the innermost metallic tubes in the bundle.

velopment, can be very competitive with the corresponding Si devices. Further refinements can be expected by additional reductions in t_{ox} and the use of high-ε insulators. We have already seen significant improvements by using HfO_2 as a gate insulator [32], [52]. An insight to the ultimate potential of CNTFETs is provided by a recent study [33]. In this nanotube FET, the role of the gate was played by a droplet of an electrolyte connected to an electrochemical electrode. The combination of an electrolyte dielectric constant of about 80 and of the ultrathin (~0.5 nm) Hemholtz layer in the electrolyte led to an extremely high transconductance of about 20 μS. Further opportunities for improvements in the CNTFET performance arise from new insights on the switching mechanism in the CNTFET, as will be discussed in Section V.

IV. Transistors From Nanotube Bundles: Bundle Composition Engineering

A major impediment to the large-scale fabrication of CNTFETs is the fact that the current synthetic schemes for SWCNTs generate mixtures of metallic (m) and semiconducting (s) nanotubes. These tubes tend to adhere to each other, forming "bundles," or "ropes" [34]. No good methods exist for the preparation of only m- or s-CNTs by selective synthesis or postsynthesis separation. If CNTFETs were to be fabricated from such a bundle, the m-CNTs in the bundle would short out the device, as shown in Fig. 6. Currently, dilute suspensions of CNT bundles are ultrasonicated and dispersed on a wafer. AFM imaging is then used to identify isolated CNTs to build the CNTFET. We have developed

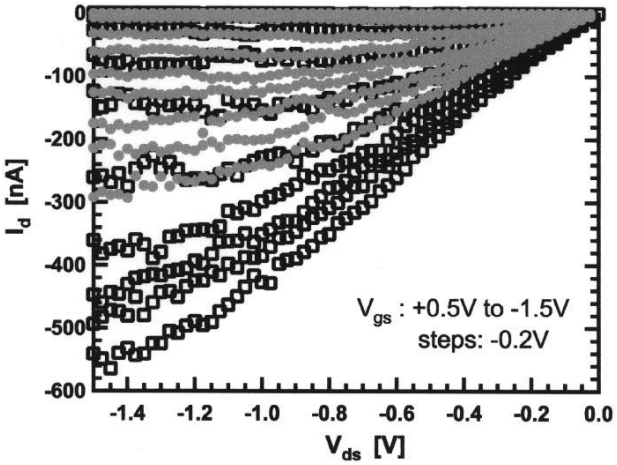


Fig. 7. Drain current versus gate voltage curves of a CNTFET upon interchange of its source and drain [32], [52].

the technique of "constructive destruction" [20] that allows us to selectively destroy the m-CNTs in a rope, leaving the s-CNTs intact.

As we discussed earlier, CNTs can carry enormous current densities at low electron energies. At higher energies, however, optical phonon excitation is possible [35], [36]. This leads to current saturation and the deposition of large amounts of energy, which eventually destroys the CNT structure [36]. To apply this method to remove m-CNTs from bundles, we first deposit the bundles on an oxidized Si wafer, then we fabricate on them an array of source drain and side gate electrodes. By applying an appropriate voltage bias to the gate, the s-CNTs can be depleted of their carriers. Then when a sufficiently high source-drain bias V_{ds} is applied, the generated current passes only through the m-CNTs, leading to their destruction, while leaving the s-CNTs essentially intact (see Fig. 6) [20]. In this way arrays of CNTFETs can be generated [20].

V. The Switching Mechanism of Carbon Nanotube Transistors

Up to this point we have implicitly assumed that the CNTFET switching mechanism is the same as that of conventional silicon devices. However, a number of observations suggest otherwise. For example, Fig. 7 shows the I_d versus V_{ds} curves for the same CNTFET upon interchange of the source and drain [32], [52]. A different current is obtained in the two cases. If the operation of the device were to be dictated by the properties of the bulk CNT, then the saturation current would be the same, since both sets of curves are taken with the same CNT. However, the different characteristics can be accounted for if transport in the tube is dominated by barriers (Schottky barriers[1]) at the CNT–metal contacts. In that case a small asymmetry of the barriers at the source and drain junctions could result in different saturation

[1]We use the term Schottky barriers in its most general sense to account for band bending in a semiconductor at a metal/semiconductor interface. When arguing about the switching in CNT transistors, we assume that there is no additional barrier present in our devices.

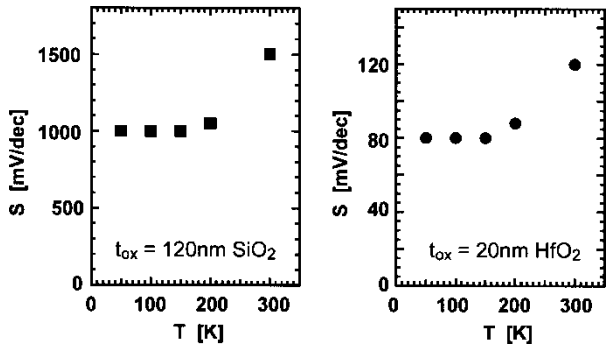


Fig. 8. Temperature dependence of the inverse subthreshold slopes, S, of two CNTFETs with 120 nm of SiO_2 and 20 nm of HfO_2 gate oxides, respectively [32], [52].

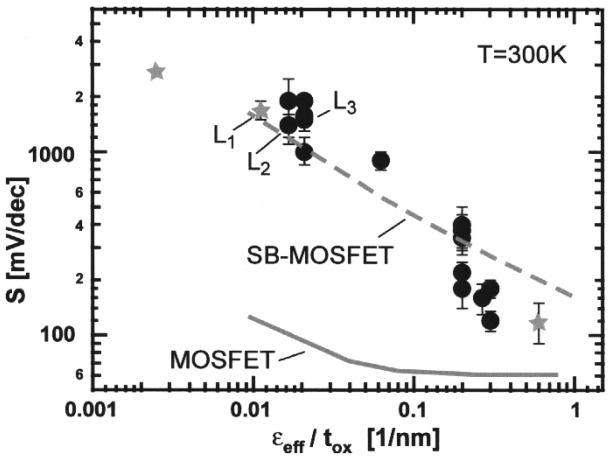


Fig. 9. Dependence of the inverse subthreshold slope S on the ratio ε_{eff}/t_{ox}, where ε_{eff} is the effective dielectric constant and t_{ox} is the thickness of the gate oxide. The points are experimental data from our own work, as well as from the literature. The solid line is the predicted behavior for bulk switching, while the dashed line is the prediction for the Schottky barrier switching model [32], [52].

currents. Further support for the existence of these barriers comes from the study of the subthreshold characteristics of CNTFETs.

In long-channel FET devices, the drain current I_D varies exponentially with V_G, and for drain bias $V_D > 3\,k_B T/q$ it is essentially independent of V_D [37]. A device characteristic of particular importance is the gate-voltage swing, or inverse subthreshold slope S. This is given by $S = \ln 10 \cdot dV_G/d(\ln I_D) \simeq (k_B T/q) \cdot \ln 10(1 + C_D/C_G)$. For a fully depleted device, the depletion capacitance C_D is zero; therefore, the second term in parenthesis becomes one. CNTs are a perfect example of a device exhibiting $C_D = 0$, since no charge variation can occur across the tube circumference. Under these ideal conditions S depends only on the temperature and has a value of 60 mV/dec at 300 K. Deviations appear when interface trap states are present in the oxide [37]. The capacitance due to these interface states $C_{\rm int}$ is in parallel with the depletion capacitance. The early back-gated CNTFETs with thick (100–150 nm) gate oxides had unexpectantly high S values of 1–2 V/dec. Devices with thin oxides, such as the top-gated CNTFET in Fig. 5, have S$\sim$100–150 mV/dec. Interface traps could in principle account for these observations. However, we found that the high S values for thick oxides are very similar independent of whether an n-type, a p-type, or an ambipolar transistor was measured. Since there is no doubt about the presence of Schottky barriers in case of an ambipolar device, and—as we proved [32], [52]—a Schottky barrier model alone can explain the trend of S as a function of $t_{\rm ox}$ quantitatively, interface traps are not responsible for the high S values in case of thick gate oxides. Further evidence that S is determined by Schottky barriers comes from its temperature dependence. As can be seen from Fig. 8, S is temperature dependent at higher temperatures but levels off at temperatures below about 200 K, suggesting a carrier tunneling process. The described subthreshold behavior may be unexpected for a bulk-switched device, but can be fully accounted for by a Schottky barrier transistor model [32], [52]. Calculations of S as a function of $t_{\rm ox}$ based on the SB and bulk models along with our own and literature experimental data are shown in Fig. 9. The data are fitted well by the SB model but not by the bulk switching model.

Focusing on the Schottky barrier itself, we note that the electrostatics in 1-D is different than that in 3-D and it is reflected in the shape of the barriers [38]. In 1-D the barriers are thin, showing an initial sharp drop followed by a logarithmically decreasing tail. Because of the thinness of the barrier, tunneling in 1-D is easy and can dominate transport. In Fig. 10(a) we show the schematic of a top-gated CNTFET and the electric field generated by the gate in that structure. Fig. 10(b) gives the conductance of the device as a function of the gate voltage for different values of the Schottky barrier height. Finally, Fig. 10(c) shows the source Schottky barrier for a mid-gap CNT at three different gate voltages [38]. The shape of the barrier and the thinning resulting from the gate field can be clearly seen.

The SB model can also provide an explanation for the long-standing problem involving the effect of the ambient on the performance of CNTFETs. Already in the first studies of CNTFETs, it was observed that although the CNTs used were not intentionally doped, the fabricated CNTFETs were p-type [24], [25]. These early FETs were fabricated in air. It was later found that when CNTs are placed in vacuum, their electrical resistance increased and their thermopower changed sign [40]. It was proposed that the CNTs transfer electrons to atmospheric O_2 and, thus, become doped with holes [41].

We have performed detailed studies of this gas effect on nanotubes in a CNTFET configuration [42]. In our early studies, we used a back-gated FET configuration with a thick (100–150 nm) gate oxide. As Fig. 11 shows, initially the air-exposed CNTFET was p-type. However, after annealing under vacuum becomes n-type. Furthermore, as Fig. 11 shows this unexpected transformation is reversible; reexposure to O_2 brings back the p character of the FET. Intermediate stages where the CNTFET exemplifies *ambipolar* behavior are clearly seen.

The above findings prove that the p character of the CNTFET is not an intrinsic property of the CNTs, but results

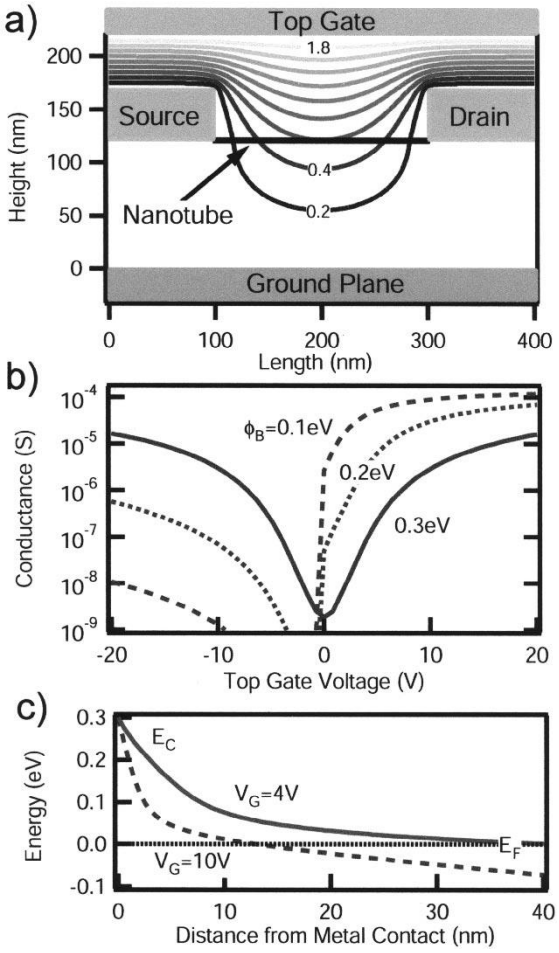


Fig. 10. Simulation of the operation of a top-gated Schottky barrier CNTFET. (a) The structure of the CNTFET and the electric field lines ($v_{gate} = 2$ V).(b) Conductance versus gate voltage for different values of the Schottky barrier. (c) The effect of gate bias on the Schottky barrier at the source–CNT junction [39].

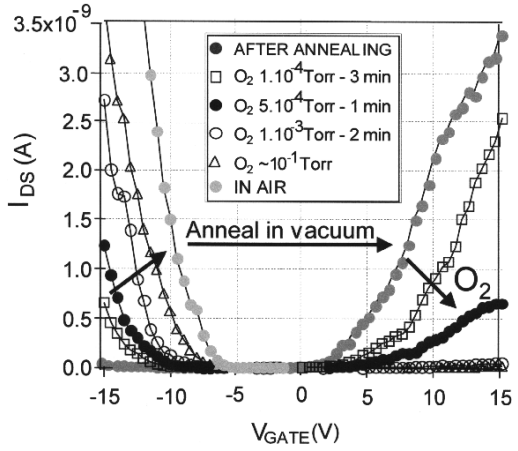


Fig. 11. Transformation of a p-type CNTFET by annealing in vacuum into an n-type CNTFET and the reverse transformation upon exposure to O_2. The transformation proceeds via ambipolar intermediate states of the CNTFET. No threshold shifts are observed, and the drain current at $v_{GATE} = 0$ does not change.

from the interaction with O_2. We have performed a number of experiments to ascertain the nature of this interaction. In Fig. 12 we show the behavior of a CNTFET upon doping

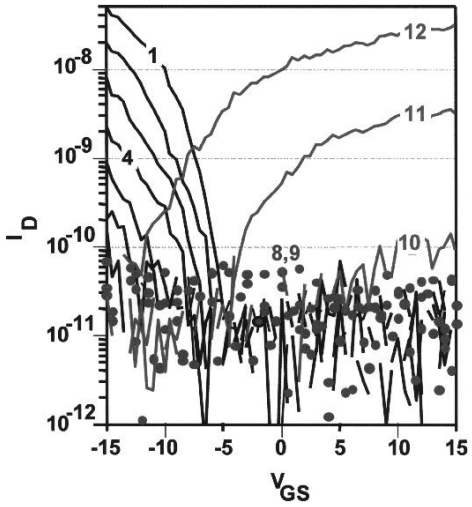


Fig. 12. The effect of doping with increasing amounts of potassium on the electrical characteristics of an initially p-type (curve 1) CNTFET.

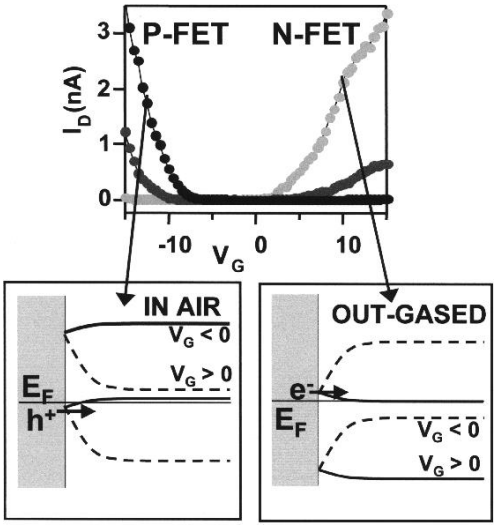


Fig. 13. Qualitative diagram showing the lineup of the valence and conduction bands of a CNT with the metal Fermi level at the source–CNT junction first in air and after annealing in vacuum.

with an electron donor, in this case potassium [42]. The well-known characteristics of doping, i.e., a shift of the threshold gate voltage and an increasing current at $V_g = 0$, are clearly seen. However, this behavior is in stark contrast with the behavior observed upon O_2 exposure (see Fig. 11). From this and other experiments, we have concluded that although some doping by O_2 may take place, this by itself cannot account for the observed behavior.

As we have already shown above, transport and switching in CNTFETs is controlled by the Schottky barriers at the contacts. Oxygen then must affect these barriers. When a nanotube is bonded to a metal electrode the resulting charge transfer determines the lineup of the nanotube bands. However, this charge transfer and the resulting field can be strongly affected by the coadsorption of other species such as oxygen near or at the CNT–metal junction. These coadsorbates can change locally the surface potential [39] or

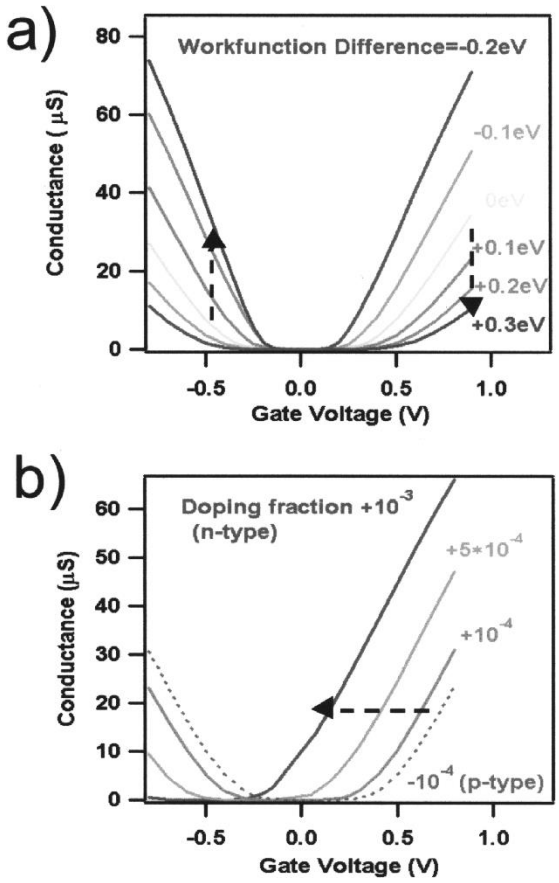


Fig. 14. Simulations of the effect of O and K on the CNTFETs current–voltage characteristics [39].

directly interact with the junction [43]. In Fig. 13 we give a schematic that accounts qualitatively for the band lineup at a metal–CNT junction in air and after annealing in a vacuum. The p character in air is the result of Fermi level pinning near the valence band maximum.

Support for this interpretation is provided by theoretical modeling. In order to take advantage of the simpler electrostatics, we used a model of a CNT surrounded by a cylindrical gate. In Fig. 14 we simulate the I–V characteristics of a CNTFET upon oxygen adsorption by changing the local surface potential [39]. Although the numbers cannot be compared because of the different device geometry used in the calculation, it is clear that the resulting behavior is similar to that of Fig. 11. In contrast, a model where fixed charges are placed along the length of the tube in order to simulate a doping interaction gives a behavior similar to that of doping by potassium in Fig. 12.

The effect of oxygen dominates the behavior of CNTFETs as long as the gate field is weak. This was the case for our earlier structures with thick gate oxides. However, in our recent CNTFETs with thin oxides ($t_{\rm ox}$ ~2–5 nm), we observe ambipolar characteristics even in air, indicating a near midgap lineup of the CNT bands [53]

VI. Multiwalled Nanotube Field-Effect Transistors

Our discussion above has been limited to transistors made out of SWCNTs. Low-temperature studies of the

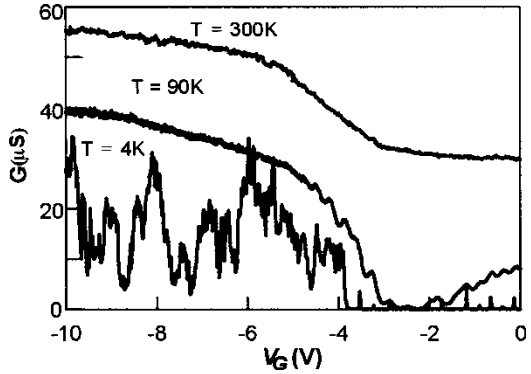


Fig. 15. Temperature dependence of the conductance G of an MWCNT as a function of the gate voltage.

Aharonov–Bohm effect in MWCNTs have concluded that in MWCNTs side-bonded to metal electrodes, effectively only the outer shell contributes to electrical transport [22]. One would expect then that MWCNTs with a semiconducting outer shell could be used to fabricate CNTFETs. However, in semiconducting CNTs, the band gap ($E_{\rm Gap}$) is inversely proportional to the tube diameter; therefore, only small-diameter MWCNTs are expected to display large switching ratios at room temperature.

In Fig. 15 we show the effect of the gate voltage V_G on the conductance G of a 14-nm diameter MWCNT, at different temperatures [21]. Clearly, the gate can modulate the conductance of the MWCNT channel at room temperature. However, there is a large residual conductance, which can be assigned to the coupling of the outer semiconducting shell to an inner metallic shell. This coupling is expected to be activated with an activation energy of the order of $\sim E_{\rm Gap}/2$. At low temperatures, the contribution of the inner metallic shell is suppressed (see Fig. 15). These observations indicate that the transport characteristics of an MWCNT at ambient temperature may have contributions from more than the outer shell even for side-bonded CNTs. Because of the activated nature of the shell-to-shell transport, and given that the semiconducting gap is inversely proportional to the CNT diameter, multishell transport should be more important for large-diameter tubes.

When the applied $V_{\rm DS}$ is increased, the average energy of the carriers is expected to increase, as well as the coupling between the carbon shells. As we discussed in Section IV, higher electron energies lead to energy dissipation and breakdown. Our studies of MWCNTs have shown that initiation of the breakdown occurs at a power threshold, which is significantly lower in air than in vacuum [36]. This indicates that an "oxidation" process takes place in air. The oxidation is most likely not a purely thermal process. Calculations on graphite have shown that once certain defects are generated, e.g., di-vacancies, a self-sustaining chain reaction with oxygen can take place [44]. Because of the lower breakdown threshold in air, the shell breakdown in MWCNT proceeds sequentially from the outer to the inner shells. This shell-by-shell breakdown can be seen as a stepwise decrease in the current flowing through the CNT [see Fig. 16(a)]. It is also illustrated by AFM images of an MWCNT draped over

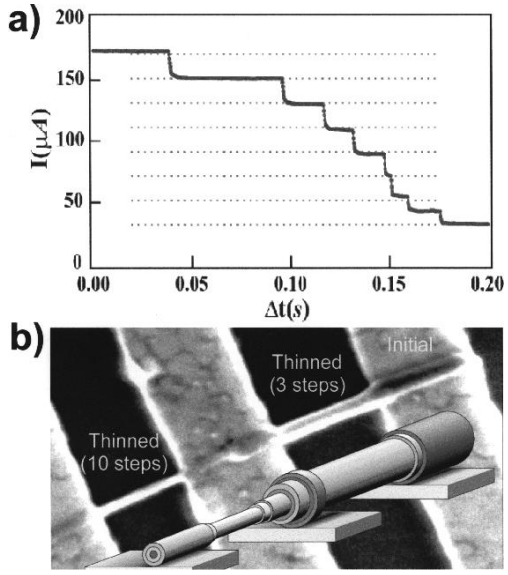


Fig. 16. (a) Stepwise decrease of the current during the breakdown of an MWCNT. (b) Images of an MWCNT showing part of the initial tube, as well as segments of the same MWCNT from which three and ten carbon atom shells have been removed.

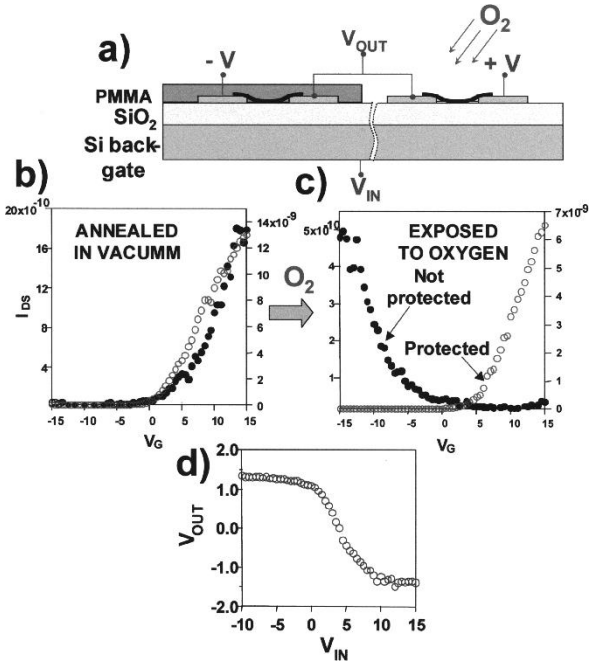


Fig. 17. (a) Schematic representation of the internanotube voltage inverter (NOT gate). One of the FETs is protected by a layer of PMMA. (b) After annealing two p-CNTFETs in vacuum to form two n-type CNTFETs. (c) After exposure to oxygen at 300 K. (d) Electrical behavior of the inverter.

several metal electrodes, so that individual CNT segments can be electrically stressed independently. Thinned segments that have lost three and ten shells, respectively, as well as a portion of the initial MWCNT, are clearly seen [Fig. 16(b)]. The capability to remove carbon atom shells one by one and identify their character through the effect of the gate field [20], [21], along with the diameter dependence of the band gap ($E_{\mathrm{Gap}} \sim 1/d_{\mathrm{CNT}}$) allows the fabrication of CNTFETs with a desired E_{Gap} can be fabricated using the controlled breakdown process.

VII. Carbon Nanotube Integrated Circuits: Logic Gates

So far we have concentrated our discussion on the performance of individual CNTFETs. The fabrication of integrated circuits using such devices is the next step. In 2001 we demonstrated that this was possible by fabricating a CMOS-like voltage inverter (a logic NOT gate) [45]. For simplicity, we used the early design of CNTFETs involving the CNT on top of gold electrodes

In Fig. 17 we show the structure and electrical characteristics of an inverter circuit involving a n- and a p-CNTFETs. Originally, both CNTFETs were p-type because of their exposure to air (oxygen). We then covered one of them by a protective film of PMMA (a more stable protection is provided by a SiO_2 film [27]), while the other was left unprotected. Both of the CNTFETs were then annealed under vacuum, which transformed both of them into n-type. After cooling, the pair was exposed to oxygen, which converted the unprotected CNTFET to p-type, while the protected one remained n-type. In this way, the two complementary CNTFETs needed were formed and wired.

The inverter works the same way as an ordinary CMOS inverter. The input voltage is applied simultaneously to the gates of the complementary CNFETs. The p-CNFET is polarized by a positive voltage, the n-FET by a negative voltage, and a common contact is used as the intermolecular inverter's output. A positive input voltage turns the n-CNFET "on" (the p-CNFET being "off"), resulting in the transmission of the negative voltage to the output. A negative input, on the other hand, turns the p-CNFET "on," producing a positive output. The electrical characteristics of the CNTFETs at each stage and those of the resulting inverter are shown in Fig. 17. We refer to this inverter circuit as an *intermolecular* inverter because it involves two nanotube molecules.

Ideally, one would like to achieve the ultimate level of integration by fabricating the circuits along the length of a single CNT, i.e., form an *intramolecular* circuit. A first realization of this approach is shown in Fig. 18(a) [44]. As the AFM image shows, the nanotube is placed on top of three prefabricated gold electrodes. In this way two back-gated initially p-type CNTFETs are formed. Then they are both covered by PMMA and a window is opened by e-beam lithography over the channel of one of them. Through this window, the channel is n-doped using potassium as a dopant. In this way two complementary FETs are formed along the same nanotube. The electrical characteristics of the resulting inverter are shown in Fig. 18(b). It is particularly interesting that despite the fact that no effort was made to optimize the construction and performance of the individual CNTFETs, the resulting inverter had a gain of almost two. This suggests that optimized CNTFETs would lead to much higher gain and can be wired along the length of a single CNT to produce more complex circuits. Following this initial work [45], more nanotube logic gates of complementary [46] or transistor–resistor [47] type have been demonstrated.

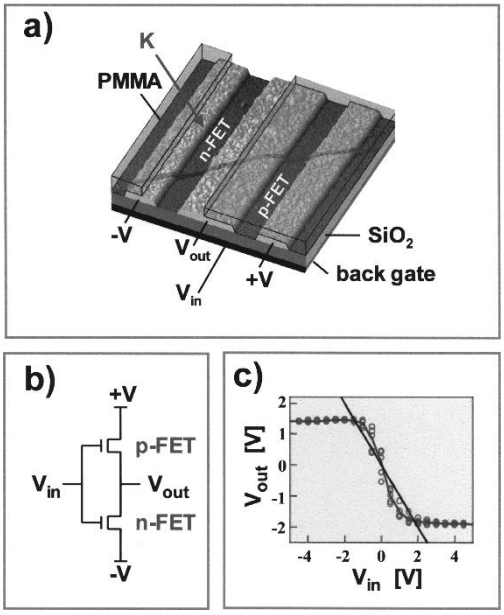


Fig. 18. (a) Atomic force microscope image of the intrananotube voltage inverter. (b) Schematic of the inverter. (c) Electrical behavior of the intrananotube inverter.

VIII. Nanotube Synthesis

SWCNTs are produced using arc discharges [2], [3], laser ablation of a carbon target [34], or CVD [48]. In all of these techniques, a metal catalyst (typically Co, Fe, and/or Ni) in the form of nanoparticles is utilized.

Currently, catalytic CVD is the most widely used technique [48]. One of the advantages of this approach is that it allows nanotubes to be grown locally by placing the catalyst at the appropriate location [49]. We have experimented with catalytic CVD using a combination of electrolytic and lithographic approaches to control both the location and orientation of the growth. The procedure shown in Fig. 19 involves the following steps:

1) patterning a thin, heavily doped silicon surface layer lithographically;
2) electrolytically etching the silicon to form porous silicon on the sidewalls of the patterned structure;
3) protecting with photoresist the areas of the silicon surface where we do not want CNT growth;
4) driving into the exposed pores the metal catalyst, and after removing the rest of the photoresist, reacting with CH_4 at 1000 C to form SWCNTs.

Fig. 20 shows SEM images of SWCNTs grown this way bridging adjacent silicon pads. These pads can subsequently be metallized.

Quite often, the presence of heavy metal catalyst particles in the nanotube product is unwanted but their removal is problematic and usually leads to damaged nanotubes. In applications in nanoelectronics, it is also likely that arrays of parallel oriented tubes will be needed in order to reduce the impedance of the devices and provide a high drive current [26]. Oriented growth of tubes is a very desirable way to achieve this type of nanotube organization. Postsynthesis alignment has also been pursued [50].

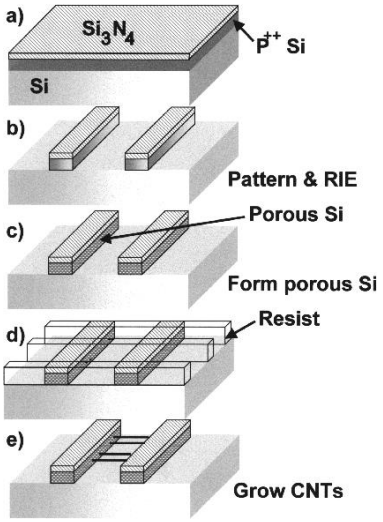


Fig. 19. Sequence of lithographic and other processing steps used to produce selective local growth of SWCNTs by chemical vapor deposition using CH_4 as the source of carbon.

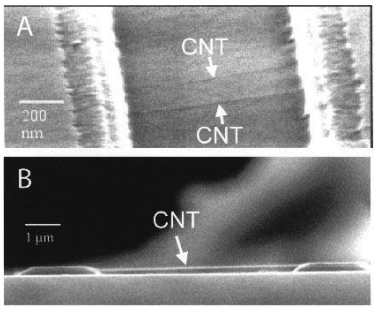


Fig. 20. Electron microscope images (top and side views) showing locally grown CNTs connecting silicon electrodes.

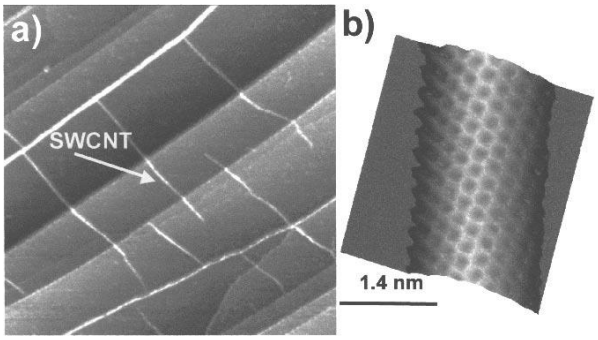


Fig. 21. (a) STM image of oriented (parallel) SWCNTs produced by heating under high vacuum to 1650 C a 6 H-SiC wafer with a (0001) Si-face. (b) Atomic resolution STM image of a SWCNT produced by this method.

Recently, we discovered a way to produce oriented SWCNTs by a catalyst-free approach [51]. This approach involves the thermal annealing under vacuum of SiC crystals. Specifically, 6H–SiC wafers with a polished (0001) Si face (this surface is equivalent to the (111) surface of cubic SiC (3 C–SiC)) were heated to 1650 C at 10^{-9} torr.

Fig. 21(a) shows a scanning tunneling microscope (STM) image of a sample cut along the $(1,-1,0,0)$ axis of the wafer resulting in a morphology characterized by well-ordered

parallel steps in the (1,1,–2,0) direction. The white 1-D structures, identified as the CNTs, extend over several steps or terraces and are not present on the samples annealed in vacuum at a temperature below 1400°C. An atomic resolution image of a semiconducting nanotube is shown in Fig. 21(b). From such images, as well as TEM images, the diameter of the nanotubes produced was determined to be in the range of 1.2–1.6 nm. These tubes are seen to have their axis perpendicular to the surface steps or be aligned along the steps. Extensive STM and AFM studies showed that this organization is uniform over their entire area of the sample. When the surface morphology is composed of terraces, the nanotubes form a weblike network with a predominance of 120°angles between straight sections (not shown) [51].

Along with individual SWCNTs, thicker tubes are also observed. Manipulation experiments using the tip of an AFM [51] indicate that these structures are SWCNT bundles, not MWCNTs.

The orientation of the nanotubes could be a result of the growth process or may involve a postsynthesis rearrangement. Our AFM experiments show that the tubes are mobile on the surface at the high temperature used for their formation. This is deduced by the observation that after perturbing the structure of the nanotubes by AFM manipulation, annealing at a temperature of 1300°C, i.e., at a temperature lower than that needed for nanotube formation, brings the manipulated nanotubes back to their parallel orientation. Thus, we believe that the orientation of the nanotubes results from their motion that releases part of the mechanical stress incorporated in the randomly grown network by favoring straight segments and by matching their orientation to the crystallographic structure of the surface. At the same time, formation of bundles contributes to the lowering of the total energy.

The above findings suggest that: 1) it may be possible to orient preformed nanotubes on an inert substrate by heating them at a temperature at which they acquire sufficient mobility and 2) in principle, it is possible to synthesize nanotubes in a controlled manner by patterning graphene strips followed by annealing. By selecting the direction of the cut of the strip, the chirality of the resulting nanotube can be chosen.

IX. Conclusion and the Future

CNTs are new materials with outstanding electrical properties. The high conductivity and exceptional stability of metallic nanotubes makes them excellent candidates for future use as interconnects in nanodevices and circuits. FETs using semiconducting CNTs have operating characteristics that are as good as or better than state-of-the-art silicon devices, and significant improvements should be expected in the near future. However, while CNTs are one of the most promising materials for molecular electronics, many challenges remain before they can become a successful technology. Most challenging are the materials issues. We still lack a method that produces a single type of CNT. In this respect, seeded growth techniques are a possibility and need to be explored. Another possible solution involves the development of efficient separation techniques, and work is pursued in this direction with encouraging initial results. The sensitivity of the electrical properties of CNTs and CNT devices to the nature of the CNT–metal contacts and the ambient environment demonstrated in this article shows that better understanding and control of these problems is absolutely essential. For CNT device integration, new fabrication techniques that are based on self-assembly of CNTs are highly desirable. While our own current interest is in computer electronics, it is likely that the initial applications of CNT devices will be in less integrated systems such as sensors, or in special applications where devices of exceptional miniaturization and performance are needed. Apart from their technological importance, CNTs are ideal model systems for the study and understanding of transport in 1-D systems and for the development of molecular fabrication technologies.

References

[1] S. Iijima, "Helical microtubules of graphitic carbon," *Nature*, vol. 354, pp. 56–58, 1991.

[2] S. Iijima and T. Ichihashi, "Single-shell carbon nanotubes of 1 nm diameter," *Nature*, vol. 363, pp. 603–605, 1993.

[3] D. S. Bethune, C. H. Kiang, M. S. Devries, G. Gorman, R. Savoy, J. Vaszquez, and R. Beyers, "Cobalt-catalyzed growth of carbon nanotubes with single atomic layer walls," *Nature*, vol. 363, pp. 605–607, 1993.

[4] M. S. Dresselhaus, G. Dresselhaus, and P. Avouris, Eds., *Carbon Nanotubes: Synthesis, Structure Properties and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[5] P. L. McEuen, M. S. Fuhrer, and H. Park, "Single-walled carbon nanotube electronics," *IEEE Trans. Nanotechnol.*, vol. 1, pp. 78–85, Mar. 2002.

[6] C. Dekker, "Carbon nanotubes as molecular quantum wires," *Phys. Today*, vol. 52, p. 22, 1999.

[7] P. G. Collins and P. Avouris, "Nanotubes for electronics," *Sci. Amer.*, vol. 283, pp. 38–45, 2000.

[8] M. S. Dresselhaus, G. Dresselhaus, and R. Saito, "Carbon fibers based on C_{60} and their symmetry," *Phys. Rev. B*, vol. 45, pp. 6234–6242, 1992.

[9] J. W. Mintmire, B. I. Dunlap, and C. T. White, "Are fullerene tubules metallic?," *Phys. Rev. Lett.*, vol. 68, pp. 631–634, 1992.

[10] S. Frank, P. Poncharal, Z. L. Wang, and W. A. deHeer, "Carbon nanotube quantum resistors," *Science*, vol. 280, pp. 1744–1746, 1998.

[11] M. Ouyang, J.-L. Huang, C. L. Cheung, and C. M. Lieber, "Energy gaps in 'metallic' single-walled carbon nanotubes," *Science*, vol. 292, pp. 702–705, 2001.

[12] J. W. Wildoer, L. C. Venema, A. G. Rinzler, R. E. Smalley, and C. Dekker, "Electronic structure of atomically-resolved carbon nanotubes," *Nature*, vol. 391, pp. 59–62, 2001.

[13] T. W. Odom, J. L. Huang, P. Kim, and C. M. Lieber, "Atomic structure and electronic properties of single-walled nanotubes," *Nature*, vol. 391, pp. 62–64, 2001.

[14] Y. Imry and R. Landauer, "Conductance viewed as transmission," *Rev. Mod. Phys.*, vol. 71, pp. S306–S312, 1999.

[15] S. Data, *Electronic Transport in Mesoscopic Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[16] P. L. McEuen, M. Bockrath, D. H. Cobden, Y.-G. Yoon, and S. G. Louie, "Disorder, pseudospins, and backscattering in carbon nanotubes," *Phys. Rev. Lett.*, vol. 83, pp. 5098–5101, 1999.
[17] L. Wenjie, M. Bockrath, D. Bozovic, J. H. Hafner, M. Tinkman, and P. Hongkum, "Fabry–Perot interference in a nanotube electron waveguide," *Nature*, vol. 411, pp. 665–669, 2001.
[18] J. Kong, E. Yenilmez, T. W. Tombler, W. Kim, H. Dai, R. B. Laughlin, L. Liu, C. S. Jayanthi, and S. Y. Wu, "Quantum interference and ballistic transmission in nanotube electron waveguides," *Phys. Rev. Lett.*, vol. 87, pp. 106 801/1–106 891/4, 2001.
[19] J. Appenzeller, R. Martel, P. Avouris, H. Stahl, and B. Lengeler, "Optimized contact configuration for the study of transport phenomena in ropes of single-wall carbon nanotubes," *Appl. Phys. Lett.*, vol. 78, pp. 3313–3315, 2001.
[20] P. G. Collins, M. S. Arnold, and P. Avouris, "Engineering carbon nanotubes using electrical breakdown," *Science*, vol. 292, pp. 706–709, 2001.
[21] P. G. Collins and P. Avouris, "Multishell conduction in multiwalled carbon nanotubes," *Appl. Phys. A*, vol. 74, pp. 329–332, 2002.
[22] A. Bachtold *et al.*, "Aharonov Bohm oscillations in carbon nanotubes," *Nature*, vol. 397, pp. 673–675, 1999.
[23] C. Schoenenberger, A. Bachtold, C. Strunk, J.-P. Salvetat, and L. Forro, "Interference and interactions in multi wall carbon nanotubes," *Appl. Phys. A*, vol. 69, pp. 283–295, 1999.
[24] S. Tans, S. Verschueren, and C. Dekker, "Room-temperature transistor based on single carbon nanotube," *Nature*, vol. 393, pp. 49–52, 1998.
[25] R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, "Single and multi-wall carbon nanotube field-effect transistors," *Appl. Phys. Lett.*, vol. 73, pp. 2447–2449, 1998.
[26] R. Martel, H.-S. P. Wong, K. Chan, and P. Avouris, "Carbon nanotube field-effect transistors for logic applications," in *Proc. IEDM 2001*, 2001, pp. 159–161.
[27] R. Martel, V. Derycke, C. Lavoie, J. Appenzeller, K. Chen, J. Tersoff, and P. Avouris, "Ambipolar electrical transport in semiconducting single-wall carbon nanotubes," *Phys. Rev. Lett.*, vol. 87, pp. 256 805/1–256 805/4, 2001.
[28] S. Wind, J. Appenzeller, R. Martel, V. Derycke, and P. Avouris, "Vertical scaling of single-wall carbon nanotubes transistors using top gate electrodes," *Appl. Phys. Lett.*, vol. 80, p. 3817, 2002.
[29] B. Yu, H. Wang, A. Joshi, Q. Xiang, E. Ibok, and M.-R. Lin, "15 nm gate length planar CMOS transistor," in *Proc. IEDM 2001*, pp. 937–939.
[30] R. Chau *et al.*, "A 50 nm depleted-substrate CMOS transistor," in *Proc. IEDM 2001*, pp. 45–48.
[31] J. Guo, S. Goasquen, M. Lundstrom, and S. Datta, "Metal-insulator-semiconductor electrostatics of carbon nanotubes," *Appl. Phys. Lett.*, vol. 81, pp. 1486–1488, 2002.
[32] J. Appenzeller, J. Knoch, V. Derycke, R. Martel, S. Wind, and P. Avouris, "Field-modulated carrier transport in carbon nanotube transistors," *Phys. Rev. Lett.*, vol. 89, pp. 126 801/1–126 801/4, 2002.
[33] S. Rosenblatt, Y. Yaish, J. Park, J. Gore, and P. L. McEuen, "High performance electrolyte gated carbon nanotube transistor," *Nano Lett.*, vol. 2, pp. 869–872, 2002.
[34] A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, X. Chunhui, L. Young Hee, K. Seong Gon, A. G. Rinzler, and D. T. Colbert *et al.*, "Crystalline ropes of metallic carbon nanotubes," *Science*, vol. 273, pp. 483–487, 1996.
[35] A. Z. Yao, C. L. Kane, and C. Dekker, "High field electrical transport in single-wall carbon nanotubes," *Phys. Rev. Lett.*, vol. 61, pp. 2941–2944, 2000.
[36] P. G. Collins, M. Hersam, M. Arnold, R. Martel, and P. Avouris, "Current saturation and electrical breakdown in multi-walled carbon nanotubes," *Phys. Rev. Lett.*, vol. 86, pp. 3128–3131, 2001.
[37] S. M. Sze, *Physics of Semiconducting Devices*. New York: Wiley, 1981.
[38] F. Leonard and J. Tersoff, "Novel length scales in nanotube devices," *Phys. Rev. Lett.*, vol. 84, pp. 4693–4696, 2000.
[39] S. Heinze, J. Tersoff, R. Martel, V. Derycke, J. Appenzeller, and P. Avouris, "Carbon nanotubes as Schottky barrier transistors," *Phys. Rev. Lett.*, vol. 89, pp. 106 801/1–106 801/4, 2002.
[40] P. G. Collins, K. Bradley, M. Ishigami, and A. Zettl, "Extreme oxygen sensitivity of electronic properties of carbon nanotubes," *Science*, vol. 287, p. 1801, 2000.
[41] S.-H. Jhi, S. G. Louie, and M. L. Cohen, "Electronic properties of oxidized carbon nanotubes," *Phys. Rev. Lett.*, vol. 85, pp. 1710–1713, 2000.
[42] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Controlling doping and carrier injection in carbon nanotube transistors," *Appl. Phys. Lett.*, vol. 80, pp. 2773–2775, 2002.
[43] N. D. Lang and P. Avouris, "Effects of coadsorption on the conductance of molecular wires," *Nano Lett.*, vol. 2, pp. 1047–1050, 2002.
[44] S. M. Lee, Y. H. Lee, Y. G. Hwang, J. R. Hahn, and H. Kang, *Phys. Rev. Lett.*, vol. 82, pp. 217–220, 1999.
[45] V. Derycke, R. Martel, J. Appenzeller, and P. Avouris, "Carbon nanotube inter- and intra-molecular logic gates," *Nano Lett.*, vol. 1, pp. 453–456, 2001.
[46] A. Javey, Q. Wang, A. Urai, Y. Li, and H. Dai, "Carbon nanotube transistor arrays for multi-stage complementary logic and ring oscillators," *Nano Lett.*, vol. 2, pp. 929–932, 2002.
[47] A. Bachtold, P. Hadley, T. Nakanishi, and C. Dekker, "Logic circuits with carbon nanotube transistors," *Science*, vol. 294, pp. 1317–1320, 2001.
[48] N. Franklin and H. Dai, "An enhanced CVD approach to extensive nanotube networks with directionality," *Adv. Mater.*, vol. 12, pp. 890–894, 2000.
[49] A. Cassel, N. Franklin, T. Tombler, E. Chan, J. Han, and H. Dai, "Current directed growth of free-standing single-walled carbon nanotubes," *J. Amer. Chem. Soc.*, vol. 121, pp. 7975–7976, 1999.
[50] Y. Zhang, A. Chang, J. Cao, Q. Wang, W. Kim, Y. Li, N. Morris, E. Yenilmez, J. Kong, and H. Dai, "Electric-field-directed growth of aligned single-walled carbon nanotubes," *Appl. Phys. Lett.*, vol. 79, pp. 3155–3157, 2001.
[51] V. Derycke, R. Martel, M. Radosavljevic, F. M. Ross, and P. Avouris, "Catalyst-free growth of ordered single-walled carbon nanotube networks," *Nano Lett.*, vol. 2, pp. 1043–1046, 2002.
[52] J. Appenzeller, J. Knoch, V. Derycke, R. Martel, S. Wind, and P. Avouris, "Short-channel like effects in Schottky barrier carbon nanotube field-effect transistors," in *Proc. IEDM 2002*, pp. 285–288.
[53] M. Radosavljevic, S. Heinze, J. Tersoff, and P. Avouris, "Drain voltage scaling in carbon nanotubes," *Appl. Phys. Lett.*, vol. 83, pp. 2435–2437, 2003.

ABOUT MRGI

Malla Reddy group of Institutions is one of the biggest conglomerates of hi-tech professional educational institutions in the state of Telangana, established in 2001 sprawling over 200 acres of land. The group is dedicated to impart quality professional education like pharmacy, Engineering & Technology, MCA, MBA courses. Our sole objective is to turn out high caliber professionals from those students who join us.

ABOUT MRCE

Malla Reddy group of Engineering (Formerly CM Engineering College) has been established under the aegis of the Malla Reddy Group of Institutions in the year 2005, a majestic empire, founded by chairman Sri Ch.Malla Reddy Garu. He has been in the field of education for the last 23 years with the intention of spearheading quality education among children from the school level itself. Malla Reddy College of Engineering has been laid upon a very strong foundation and has ever since been excelling in every aspect. The bricks of this able institute are certainly the adept management, the experienced faculty, the selfless non-teaching staff and of course the students.

ABOUT ICTIMES

ICTIMES started long back with its banner to promote the vision of future technologies that change the trends of life on this planet earth. Under this banner, the Department of Electronics and Communication Engineering at MRCE organizes the ICLTEC – International Conference on Latest Trends in Electronics and Communication to provide a scholarly platform to ignite the spirit of Research and bring out the latent potential in teaching fraternity and student community. ICLTEC accommodates major areas like, signal processing, Embedded systems, Network Security, Green Technology, Digital signal processing, Communication systems.

ABOUT ICLTEC

International Conference on Latest Trends in Electronics and Communication **(ICLTEC-2019)** will bring together innovative academicians, researchers and industrial experts in the field of Electronics & Communication to a common forum. The idea of the conference is, for the scientists, scholars, engineers and students from the Universities across the world and the industry as well, to present ongoing research activities, and hence of foster research relations between the Universities and the industry with the rapid development of trends and studies in the fields concerned.
will provide a heartwarming platform to researchers, scholars, faculty and students to exchange their novel ideas face to face together.

Estd :2005

MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE - New Delhi, Affiliated to JNTU - Hyderabad,Accredited by NBA & Accredited by NAAC.
ISO 9001:2015 Certified Institution, Recognition of College under Section 2(f) & 12 (B) of the UGC Act,1956.
Address: Maisammaguda, Dhulapally,(Post Via Kompally),Secunderabad - 500 100.
Ph: 040-64632248, 9348161222 ,9346162620. Email: principal@mrce.in
Website : www.mrce.in